CAPACITY-ACHIEVING CODING MECHANISMS: SPATIAL COUPLING AND GROUP SYMMETRIES

A Dissertation

by

SANTHOSH KUMAR VANAPARTHY

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jean-Francois Chamberland
Co-Chair of Committee,	Henry D. Pfister
Committee Members,	Srinivas Shakkottai
	Matthew A. Papanikolas
Head of Department,	Miroslav M. Begovic

December 2015

Major Subject: Electrical Engineering

ABSTRACT

The broad theme of this work is in constructing optimal transmission mechanisms for a wide variety of communication systems. In particular, this dissertation provides a proof of threshold saturation for spatially-coupled codes, low-complexity capacity-achieving coding schemes for side-information problems, a proof that Reed-Muller and primitive narrow-sense BCH codes achieve capacity on erasure channels, and a mathematical framework to design delay sensitive communication systems.

Spatially-coupled codes are a class of codes on graphs that are shown to achieve capacity universally over binary symmetric memoryless channels (BMS) under belief-propagation decoder. The underlying phenomenon behind spatial coupling, known as "threshold saturation via spatial coupling", turns out to be general and this technique has been applied to a wide variety of systems. In this work, a proof of the threshold saturation phenomenon is provided for irregular low-density parity-check (LDPC) and low-density generator-matrix (LDGM) ensembles on BMS channels. This proof is far simpler than published alternative proofs and it remains as the only technique to handle irregular and LDGM codes. Also, low-complexity capacity-achieving codes are constructed for three coding problems via spatial coupling: 1) rate distortion with side-information, 2) channel coding with side-information, and 3) write-once memory system. All these schemes are based on spatially coupling compound LDGM/LDPC ensembles.

Reed-Muller and Bose-Chaudhuri-Hocquengham (BCH) are well-known algebraic codes introduced more than 50 years ago. While these codes are studied extensively in the literature it wasn't known whether these codes achieve capacity. This work introduces a technique to show that Reed-Muller and primitive narrow-sense BCH codes achieve capacity on erasure channels under maximum a posteriori (MAP) decoding. Instead of relying on the weight enumerators or other precise details of these codes, this technique requires that these codes have highly symmetric permutation groups. In fact, any sequence of linear codes with increasing blocklengths whose rates converge to a number between 0 and 1, and whose permutation groups are doubly transitive achieve capacity on erasure channels under bit-MAP decoding. This provides a rare example in information theory where symmetry alone is sufficient to

achieve capacity.

While the channel capacity provides a useful benchmark for practical design, communication systems of the day also demand small latency and other link layer metrics. Such delay sensitive communication systems are studied in this work, where a mathematical framework is developed to provide insights into the optimal design of these systems.

To my Teachers

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

The fundamental goal of communication is to reproduce information at one entity (known as the receiver) that is held by another (known as the transmitter) via a permeable medium (known as the channel). The information is represented by a set of symbols agreed on beforehand between the entities. Such a communication is associated with a *rate* defined as the number of symbols transmitted in a given unit time. Usually, the communication medium introduces errors that can cause some confusion between symbols at the receiver, preventing the perfect reliability of the information. Until the seminal work of Shannon [1], it was believed that arbitrarily high reliability can only be achieved at the expense of arbitrarily small communication rate. Shannon introduced a mathematical quantity called *channel capacity*, a theoretical limit on the rate below which communication is possible with arbitrary reliability, and above which it is not. To achieve reliable communication below the channel capacity, Shannon used an ingenious random coding mechanism. While this scheme is appealing theoretically, it is infeasible to implement this scheme in practice due to its computational complexity.

While coding theory itself predates Shannon's work, the introduction of channel capacity has brought a new challenge to theorists and practitioners alike: to approach the fundamental capacity limit with low computational complexity. The focus in coding theory for several decades after the birth of information theory has largely been on the so-called linear codes. Binary linear codes are defined as the k-dimensional subspace of the n-dimensional vector space $\{0,1\}^n$ over the binary field $\{0,1\}$. Such a code is said to be an (n,k) binary linear code and it is associated with a rate of k/n. The earliest instance of this class, the Hamming code, was developed before Shannon's work. An important parameter of these codes is the *minimum distance*, which is defined as the minimum distance between any two codewords. An (n,k) binary linear code with a minimum distance of $d_{\min}$ can correct $\lfloor \frac{d_{\min}-1}{2} \rfloor$ errors. Thus, codes with large minimum distance can correct large number of errors. Much of early work in coding theory focused on constructing linear codes with large $d_{\min}$ for a given (n,k). Several elegant codes were discovered. A few pop-

ular codes include Reed-Muller codes, Bose-Chaudhuri-Hocquengham (BCH) codes, Reed-Solomon codes, convolutional codes.

A sequence of (n, k) linear codes is said to be *asymptotically good* if the sequence of code rates k/n and normalized minimum distances d/n are bounded away from 0. Asymptotically good codes are desirable because they correct a constant fraction of errors while simultaneously having a non-negligible rate. Unfortunately, none of the codes above, Reed-Muller, BCH, Reed-Solomon or convolutional are asymptotically good. During the early decades of information theory, the focus of the coding theory community was not on constructing low-complexity capacity-achieving codes but on finding codes of short block length that have good performance in practice. As such, little theoretical progress was made on determining whether these codes achieve capacity. For some classes of algebraic codes, this property is established in this dissertation.

Things changed dramatically in 1993 with the advent of Turbo codes and iterative decoding [2]. It was suddenly possible to construct codes that operate close to capacity and also have low computational complexity. In 1995, low-density parity-check (LDPC) codes were rediscovered independently by Spielman [3], Mackay and Neal [4]. LDPC codes had been introduced by Gallager in the 1960s during the early days of information theory [5]. Unfortunately, their full potential was not realized then due to the limited computing power of the time. With their rediscovery and the introduction of irregular LDPC codes, it was possible to construct capacity achieving codes with low-complexity at least for erasure channels. However, it was not until the discovery of polar codes in late 2000s that one could *provably* construct capacity-achieving codes under low complexity for general channels [6].

Polar codes are closely related to Reed-Muller codes. However, their construction is more information theoretic rather than algebraic (like Reed-Muller codes) and requires a synthesis based on the channel. Another type of codes based on graphs known as convolutional LDPC codes were introduced in 1999 [7]. During mid 2000s, terminated convolutional LDPC codes (known as spatially-coupled codes) were observed to have iterative decoding thresholds close to capacity. The underlying phenomenon behind the excellent performance turned out to be universal and spatially-coupled codes are now known to achieve the capacity of a wide variety of systems. These codes form the basis of Chapters II and III in this dissertation.

Another aspect we focus on in this dissertation is the delay sensitive communi-

cation system. Information theory generally focuses on characterizing the maximum amount of information that can be communicated by resorting to asymptotically long block lengths and consequently long delay at the receiver. However, with the increasing demand for wireless video over cellular networks, other performance metrics such as latency and quality of service are equally important. These delay sensitive communication systems are the focus of Chapter V in this dissertation. Here, we develop a mathematical framework to provide guidelines for the optimal design of delay sensitive communication systems.

An overarching theme of this dissertation lies in constructing optimal transmission mechanisms for a wide variety of communication systems. In the next section, we summarize the contributions in this dissertation.

I.A OUTLINE OF DISSERTATION

I.A.1 Threshold Saturation via Spatial Coupling

In Chapter II of this dissertation, we analyze a class of capacity achieving codes called spatially-coupled codes. These codes were introduced in [7] in 1999 as convolutional LDPC codes that combines elements from both turbo codes and convolutional codes. However, their potential was not realized immediately. Subsequently, in mid 2000s [8, 9], these codes were observed to have iterative decoding thresholds close to capacity. Recently, the reason for the excellent performance of these codes was described in [10] as *threshold saturation via spatial coupling*, and the authors therein gave a rigorous proof of the threshold saturation for the binary erasure channel.

The idea behind the construction of these codes is the following. Consider a series of LDPC ensembles and couple them locally. Then, terminate the bits at the boundary, which is similar to revealing these bits to the decoder. Under iterative decoding, the known information at the boundary propagates inward and improves the performance of the decoder even beyond the iterative decoding threshold of the original LDPC ensemble. The term threshold saturation refers to the fact that the iterative decoding threshold of the coupled ensemble will be equal to the optimum or maximum a posteriori (MAP) decoding threshold of the original LDPC ensemble. Thus, spatial coupling provides a remedy where iterative decoding falls short of MAP decoding. Spatially-coupled codes have been successfully applied in numerous areas apart from communication systems. The generality of the threshold saturation phenomenon is evident from its applicability in satisfiability problems, graph coloring,

compressed sensing.

In particular, it is now proven that spatially-coupled regular LDPC codes universally achieve capacity over the class of binary memoryless symmetric (BMS) channels under belief-propagation decoding. In 2012 [11, 12], potential functions have been used to simplify threshold saturation proofs for scalar and vector recursions. Our contribution in Chapter II is in proving the threshold saturation phenomenon for general BMS channels using potential functions. Our method yields a far simpler proof compared to the previous proof and it remains as the only existing method that handles irregular ensembles and low-density generator matrix (LDGM) ensembles. The potential function approach we use is inspired by physics and blends naturally with the philosophy of spatial coupling. The required potential functions are closely related to the average Bethe free entropy of the ensembles in the large-system limit. These functions also appear in statistical physics when the replica method is used to analyze optimal decoding.

I.A.2 Spatially-Coupled Codes for Side-Information Problems

In Chapter III, we continue the theme of spatial coupling from Chapter II. In particular, we construct low-complexity capacity achieving coding schemes based on spatial coupling for three problems: 1) rate distortion with side information (the Wyner-Ziv formulation) 2) channel coding with side information (the Gelfand-Pinsker formulation) 3) a write-once memory system. In all cases, we consider systems with binary symmetric sources and channels. The coding scheme for all these problems is based on compound LDGM/LDPC codes. Compound LDGM/LDPC codes were shown to achieve the capacity of binary instances of Wyner-Ziv and Gelfand-Pinsker problems under MAP decoding [13]. Our application of the compound codes to the write-once memory (WOM) system is new. The coset decomposition in the compound LDGM/LDPC codes naturally provides a way to simultaneously encode the required messages and provide error protection. Even though the MAP decoding of these codes achieves capacity, iterative decoding for these problems falls far short. The main result in this chapter is that spatial coupling enables the compound codes to achieve the capacity region of these problems with low-complexity message-passing encoding and decoding algorithms.

We note that standard BP algorithms do not yield desired performance for problems involving source coding. For lossy compression, one requires a crucial modifica-

tion to the BP algorithm known as guided decimation. In the belief propagation with guided decimation (BPGD) algorithm, for every few iterations of the BP algorithm a bit with largest bias is fixed to match the bias. This guides the BP algorithm to converge to a codeword. Spatially-coupled LDGM ensembles with the BPGD algorithm were observed to achieve the rate distortion limit. We observe this to be the case even with compound LDGM/LDPC codes. There is an additional complication with decimation for the compound codes. The decimated bits are required to satisfy the parity constraints to yield a valid codeword. Our simulations never yielded a valid codeword when the BPGD algorithm was applied (without spatial-coupling) to compound LDGM/LDPC codes. However, spatial coupling compound LDGM/LDPC codes with the proper termination conditions enables the BPGD algorithm to find a valid codeword.

For the WOM problem, we consider both noiseless systems and systems with read errors. The encoding for this problem with compound LDGM/LDPC codes is an instance of the erasure quantization. This reduction to the erasure quantization allows an efficient linear complexity encoding algorithm.

I.A.3 Capacity-Achieving Codes via Group Symmetry

In Chapter IV, we introduce a new approach to proving that a sequence of deterministic linear codes achieves capacity. This approach is valid on an erasure channel under MAP decoding. Rather than relying on the precise structure of the codes, our method requires only that the codes are highly symmetric. In particular, the technique applies to any sequence of linear codes where the blocklengths are strictly increasing, the code rates converge to a number between 0 and 1, and the permutation group of each code is doubly transitive. This also provides a rare example in information theory where symmetry alone implies near-optimal performance.

An important consequence of this result is that a sequence of Reed-Muller codes with increasing blocklength achieves capacity if its code rate converges to a number between 0 and 1 [14]. This possibility has been suggested previously in the literature but it has only been proven for cases where the limiting code rate is 0 or 1. Moreover, these results extend naturally to affine-invariant codes and, thus, to all extended primitive narrow-sense BCH codes. Our proof is based on the analysis of extrinsic information transfer (EXIT) functions for linear codes. For erasure channels, it is possible to characterize EXIT functions in terms of monotone boolean functions. For

a sequence of linear codes to be capacity achieving, the sequence of EXIT functions must converge to a 0-1 step function and due to the so-called *area theorem* for these functions, this transition point must be at the capacity limit. It is possible to show such a threshold behavior for these EXIT functions when the code satisfies certain symmetry conditions.

Since Reed-Muller and BCH codes are not asymptotically good, they cannot correct all patterns with a constant fraction of erasures and simultaneously have non-vanishing rate. However, to achieve capacity on erasure channels, it is sufficient to correct almost all patterns of erasures up to the capacity limit, which is the case for Reed-Muller and some BCH codes. These results are rather surprising. Until polar codes were discovered, it was believed that achieving capacity may require some degree of randomness [15–17]. While polar codes are purely deterministic, they require a synthesis that is highly dependent on the channel. Thus, the capacity achieving nature of polar codes seems unrelated to the symmetry. On the other hand, our approach guarantees that a doubly transitive permutation group is sufficient to achieve capacity.

This result also shows for the first time that there exists a sequence of cyclic codes that achieves capacity on erasure channel. This is due to the fact that all BCH codes and punctured Reed-Muller codes are cyclic codes. While our proof is heavily dependent on the structure of the erasure channel, several aspects of it can be generalized using generalized EXIT (GEXIT) functions. Nevertheless, a proof for the general BMS channels is not yet known.

I.A.4 Code-Rate Selection via Hitting Time and Large-Deviations

In Chapter V, we consider the performance of delay sensitive digital communication systems. We develop a mathematical framework that provides insight into the optimal allocation of link layer resources, primarily code rate of the communication under the delay constraints. Messages are transmitted over finite-state erasure channels with memory, and the information is protected using error-correcting codes; successful receptions of codewords are acknowledged at the source through instantaneous feedback. A distinguishing feature of this framework is the rigorous modeling of the bit erasure channel and a comprehensive treatment of channels with memory, error control coding and queuing. The primary focus of this work is on delay-sensitive applications, codes with finite block lengths and, necessarily, non-vanishing proba-

bilities of decoding failure.

The contribution of our work is twofold. A methodology based on generating functions is introduced to compute the distribution of the time required to empty a buffer. Based on this distribution, the mean hitting time to an empty queue and delay-violation probabilities for specific thresholds can be computed explicitly. The proposed techniques apply to situations where the transmit buffer contains a predetermined number of information bits at the onset of the data transfer. Focusing on the hitting time implicitly provides the distribution of the time an incoming delay packet faces. This viewpoint also offers a foundation for choosing among possible routes and interfaces. This also obviates the search for representative arrival processes. The analysis works well for relatively small buffer sizes. Our study differs from previous contributions in the literature without resorting to asymptotically long coding delays or approximations.

Under large buffer sizes, our technique based on generating functions becomes cumbersome. For this, large deviation principles governing the system evolution provide meaningful guidelines for resource allocation. In particular, large-deviations are obtained for the empirical mean service time and the average packet-transmission time associated with the communication process. The argument of the rate function in the large deviations can be selected to adjust the delay-sensitivity to the needs of the underlying data flow. This rigorous framework yields a pragmatic methodology to select code rate and block length for the communication unit as functions of the service requirements. Examples motivated by practical systems are provided to further illustrate the application of these techniques.

CHAPTER II

THRESHOLD SATURATION FOR SPATIALLY-COUPLED LDPC AND LDGM CODES ON BMS CHANNELS*

II.A INTRODUCTION

Low-density parity-check (LDPC) convolutional codes were introduced in [7] and shown to have outstanding performance under belief-propagation (BP) decoding in [8, 9, 18]. The fundamental principle behind this phenomenon is described by Kudekar, Richardson, and Urbanke in [10] and coined *threshold saturation via spatial coupling*. Roughly speaking, multiple LDPC ensembles are placed next to each other, locally coupled together, and then terminated at the boundaries. The number of LDPC ensembles is called the *chain length* and the range of local coupling is determined by the *coupling width*. This termination at the boundary can be regarded as perfect side information for decoding. Under iterative decoding, this "perfect" information propagates inward and dramatically improves performance. See [10] for a rigorous construction of spatially-coupled codes, and [19] for a comprehensive discussion of these codes.

For the binary erasure channel (BEC), spatially coupling a collection of (d_v, d_c)-regular LDPC ensembles produces a new ensemble that is nearly regular. Moreover, the BP threshold of the coupled ensemble approaches the maximum a posteriori (MAP) threshold of the original ensemble [10]. Recently, a proof of saturation to the *area threshold* has been given for (d_v, d_c)-regular LDPC ensembles on binary memoryless symmetric (BMS) channels under mild conditions [19]. This result implies that spatially-coupled LDPC codes achieve capacity *universally* over the class of BMS channels under iterative decoding because the area threshold of regular LDPC codes can approach the Shannon limit uniformly over this class. Here, universality refers to the notion that the same ensemble works well for the entire class of BMS channels [19, Lemma 29].

The idea of threshold saturation via spatial coupling has started a small revolution in coding theory, and spatially-coupled codes have now been observed to approach the capacity regions of many systems [18, 20–25]. For spatially-coupled systems with suboptimal component decoders, such as message-passing decoding of code-division multiple access (CDMA) [26, 27] or iterative hard-decision decoding of spatially-coupled generalized LDPC codes [28], the threshold saturates instead to an intrinsic threshold defined by the suboptimal component decoders.

Spatial-coupling has also led to new results for K-SAT, graph coloring, and the Curie-Weiss model in statistical physics [29–31]. For compressive sensing, spatially-coupled measurement matrices were introduced in [32], shown to give large improvements with Gaussian approximated BP reconstruction in [33], and finally proven to achieve the theoretical limit in [34]. Recent results based on spatial-coupling are now too numerous to cite thoroughly.

Recently, a simple approach, based on potential functions, is used in [11, 12] to prove that the BP threshold of spatially-coupled irregular LDPC ensembles over a BEC saturates to the conjectured MAP threshold (known as the Maxwell threshold) of the underlying irregular ensembles. This technique was motivated by [35] and is also related to the continuum approach to density evolution (DE) in which potential functions are used to prove threshold saturation for compressed sensing [34].

In this chapter, the threshold saturation proof based on potential functions in [11, 12] is extended to spatially-coupled irregular LDPC and LDGM codes on BMS channels. The main results are summarized, rather informally, in the following theorems whose proofs comprise the majority of this chapter. See the main text for precise statements and conditions under which the results hold. Moreover, for LDPC codes, we actually show threshold saturation to a quantity called the *potential threshold*. For many LDPC ensembles, it is known that the MAP threshold $\mathtt{h}^{\mathrm{MAP}}$ is upper bounded by the potential threshold. In some cases, they are actually equal (e.g., see Remark 33).

Theorem: Consider a spatially-coupled LDPC ensemble and a family of BMS channels that is ordered by degradation, and parameterized by entropy, $\mathtt{h}$. If $\mathtt{h} < \mathtt{h}^{\mathrm{MAP}}$, then, for any sufficiently large coupling width, the spatially-coupled DE converges to the perfect decoding solution. Conversely, if $\mathtt{h} > \mathtt{h}^{\mathrm{MAP}}$, then for a fixed coupling width and sufficiently large chain length, the spatially-coupled DE *does not* converge to the perfect decoding solution.

Thus, the spatially-coupled BP threshold saturates to $\mathtt{h}^{\mathrm{MAP}}$ for LDPC codes.

For LDGM codes, message-passing decoding always results in non-negligible error floors. Even when DE is initialized with perfect information, it converges to a nontrivial *minimal fixed point*. When a certain quantity, which we call the *energy gap*, is positive, the spatially-coupled DE converges to a fixed point which is elementwise better than the minimal fixed point. Also, it is conjectured that the MAP decoding performance is governed by the region where the energy gap is positive (e.g., see Section II.E.1).

Theorem: Consider a spatially-coupled LDGM ensemble and a BMS channel. If the *energy gap* for the channel is positive, then, for sufficiently large coupling width, the spatially-coupled DE converges to a fixed point which is elementwise better than the minimal fixed point of the underlying LDGM ensemble.

A variety of observations, formal proofs, and applications now bear evidence to the generality of threshold saturation. The technique in [11, 12] is based on defining a potential function. The average Bethe free entropy in the large-system limit [36,37] serves as our potential function. The crucial properties of the free entropy that we leverage are 1) stationary points of the free entropy are related to the fixed points of DE, 2) there exists a spatially-coupled potential, defined by a spatial average of the free entropy, where the fixed points of spatially-coupled DE are stationary points of the spatially-coupled potential. It is tempting to conjecture that this approach can be applied to more general graphical models by computing their average Bethe free entropy.

II.B PRELIMINARIES

II.B.1 Measures and Algebraic Structure

Any output $Y \in \overline{\mathbb{R}}$ of a binary-input communication channel, with input $X \in \{-1, +1\}$, can be represented by the log-likelihood ratio (LLR)

$$Q = \log \frac{P_{Y|X}(\alpha|1)}{P_{Y|X}(\alpha|-1)},$$

which is a sufficient statistic for X given Y. Therefore, a communication channel can be associated with a LLR distribution.

The channel is said to be *output symmetric* if

$$P_{Y|X}(\alpha|1) = P_{Y|X}(-\alpha| - 1).$$

If the channel is output symmetric, then it suffices to compute the LLR distribution conditional on $X = 1$.

Throughout this chapter, we assume the communication is over a binary-input output-symmetric memoryless channel. The distribution of the LLR is a key object in the analysis of decoding process. The distribution of the random variable Q is given by the distribution of the LLR conditional on $X = 1$. For mathematical convenience, we represent these distributions by measures on the extended real numbers $\overline{\mathbb{R}}$. Thus, Q is represented by a measure x where

$$\Pr(Q \leq t) = \mathsf{x}([-\infty, t]).$$

We call a finite signed Borel measure x on $\overline{\mathbb{R}}$ *symmetric* if

$$\mathsf{x}(-E) = \int_{-E} \mathsf{x}(d\alpha) = \int_{E} e^{-\alpha} \mathsf{x}(d\alpha),$$

for all Borel sets $E \subseteq \overline{\mathbb{R}}$, where $\overline{\mathbb{R}}$ is a compact metric space under $\tanh(\cdot)$. This necessarily implies that for any finite symmetric measure x, $\mathsf{x}(\{-\infty\}) = e^{-\infty}\mathsf{x}(\{\infty\}) = 0$. We note that binary-input output-symmetric channels give rise to symmetric measures [38, Theorem 4.27].

Equivalently, a more operational definition, a finite signed Borel measure x is symmetric if

$$\int_{-E} f(\alpha)\mathsf{x}(d\alpha) = \int_{E} f(-\alpha)e^{-\alpha}\mathsf{x}(d\alpha),$$

for all bounded measurable real-valued functions f and Borel sets $E \subseteq \overline{\mathbb{R}}$. An immediate consequence is the following Proposition.

Proposition 1: Let x be a symmetric measure and $f\colon \overline{\mathbb{R}} \to \mathbb{R}$ be an odd function that is bounded and measurable, then

$$\int f(\alpha)\mathsf{x}(d\alpha) = \int f(\alpha) \tanh\left(\tfrac{\alpha}{2}\right) \mathsf{x}(d\alpha).$$

Proof. See Section II.H.2.1. □

In particular, for a symmetric measure x and any natural number k,

$$\int \tanh\left(\tfrac{\alpha}{2}\right)^{2k-1} \mathsf{x}(d\alpha) = \int \tanh\left(\tfrac{\alpha}{2}\right)^{2k} \mathsf{x}(d\alpha).$$

This last relation is a well-known result and its utility will become apparent in the section on entropy.

Let $\mathcal{M}$ denote the set of finite signed symmetric Borel measures on the extended real numbers $\overline{\mathbb{R}}$. In this work, the primary focus is on convex combinations and differences of symmetric probability measures, which inherit many of their properties from $\mathcal{M}$. Let $\mathcal{X} \subset \mathcal{M}$ be the convex subset of symmetric probability measures. Also, let $\mathcal{X}_{\mathrm{d}} \subset \mathcal{M}$ be the subset of differences of symmetric probability measures:

$$\mathcal{X}_{\mathrm{d}} \triangleq \{\mathsf{x}_1 - \mathsf{x}_2 \mid \mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}\}.$$

In the interest of notational consistency, x is reserved for both finite signed symmetric Borel measures and symmetric probability measures, and y, z denote differences of symmetric probability measures. Also, all logarithms that appear in this chapter are *natural*, unless the base is explicitly mentioned.

In this space, there are two important binary operators, $\circledast$ and $\circledbslash$, that denote the variable-node operation and the check-node operation for LLR message distributions, respectively. Below, we give an explicit integral characterization of the operators $\circledast$ and $\circledbslash$. For $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{M}$, and any Borel set $E \subset \overline{\mathbb{R}}$, define

$$\begin{aligned}
(\mathsf{x}_1 \circledast \mathsf{x}_2)(E) &\triangleq \int \mathsf{x}_1(E - \alpha)\, \mathsf{x}_2(d\alpha), \\
(\mathsf{x}_1 \circledbslash \mathsf{x}_2)(E) &\triangleq \int \mathsf{x}_1\left(2\tanh^{-1}\left(\frac{\tanh(\frac{E}{2})}{\tanh(\frac{\alpha}{2})}\right)\right) \mathsf{x}_2(d\alpha).
\end{aligned}$$

Equivalently, for any bounded measurable real-valued function f,

$$\begin{aligned}
\int f d(\mathsf{x}_1 \circledast \mathsf{x}_2) &= \iint f(\alpha_1 + \alpha_2)\, \mathsf{x}_1(d\alpha_1)\, \mathsf{x}_2(d\alpha_2), \\
\int f d(\mathsf{x}_1 \circledbslash \mathsf{x}_2) &= \iint f(\tau^{-1}(\tau(\alpha_1)\tau(\alpha_2)))\, \mathsf{x}_1(d\alpha_1)\, \mathsf{x}_2(d\alpha_2),
\end{aligned}$$

where $\tau\colon \overline{\mathbb{R}} \to [-1,1]$, $\tau(\alpha) = \tanh\left(\frac{\alpha}{2}\right)$. Associativity, commutativity, and linearity of the operators $\circledast$, $\circledbslash$ are inherited from the underlying algebraic structure of $(\overline{\mathbb{R}}, +)$, $([-1,1], \cdot\,)$, respectively. Moreover, the space of symmetric probability measures is closed under these binary operations [38, Theorem 4.29].

In a more abstract sense, the measure space $\mathcal{M}$ along with either multiplication operator ($\circledast$, $\circledbslash$) forms a commutative monoid, and this algebraic structure is induced on the space of symmetric probability measures $\mathcal{X}$. There is also an intrinsic connection between the algebras defined by each operator and one consequence is the duality (or conservation) result in Proposition 4. The identities in these algebras, $\mathsf{e}_{\circledast} = \Delta_0$ and $\mathsf{e}_{\circledbslash} = \Delta_\infty$, also exhibit an annihilator property under the dual operation

$$\Delta_0 \circledbslash \mathsf{x} = \Delta_0, \qquad\qquad \Delta_\infty \circledast \mathsf{x} = \Delta_\infty.$$

The wildcard $*$ is used to represent either operator in statements that apply to both operations. For example, the shorthand x^{*n} is used to denote n fold operations

$$\mathsf{x}^{*n} = \underbrace{\mathsf{x} * \cdots * \mathsf{x}}_{n},$$

and this notation is extended to polynomials. In particular, for a polynomial $p(t) = \sum_{n=0}^{\deg(p)} p_n t^n$ with real coefficients, we define

$$p^*(\mathsf{x}) \triangleq \sum_{n=0}^{\deg(p)} p_n \mathsf{x}^{*n},$$

where we define $\mathsf{x}^{*0} \triangleq \mathsf{e}_*$. For the formal derivative $p'(t) = \frac{dp}{dt}$, we have

$$p'^*(\mathsf{x}) = \sum_{n=0}^{\deg(p)} n p_n \mathsf{x}^{*n-1}.$$

In general, the operators $\circledast$, $\circledbslash$ do not associate

$$\mathsf{x}_1 \circledast (\mathsf{x}_2 \circledbslash \mathsf{x}_3) \neq (\mathsf{x}_1 \circledast \mathsf{x}_2) \circledbslash \mathsf{x}_3, \qquad\qquad \mathsf{x}_1 \circledbslash (\mathsf{x}_2 \circledast \mathsf{x}_3) \neq (\mathsf{x}_1 \circledbslash \mathsf{x}_2) \circledast \mathsf{x}_3,$$

nor distribute

$$\mathsf{x}_1 \circledast (\mathsf{x}_2 \boxplus \mathsf{x}_3) \neq (\mathsf{x}_1 \circledast \mathsf{x}_2) \boxplus (\mathsf{x}_1 \circledast \mathsf{x}_3), \quad \mathsf{x}_1 \boxplus (\mathsf{x}_2 \circledast \mathsf{x}_3) \neq (\mathsf{x}_1 \boxplus \mathsf{x}_2) \circledast (\mathsf{x}_1 \boxplus \mathsf{x}_3).$$

II.B.2 Partial Ordering by Degradation

Degradation is an important concept that allows one to compare some LLR message distributions. The order imposed by degradation is indicative of relating probability measures through a communication channel [38, Definition 4.69]. The following is one of several equivalent definitions and is the most suitable for our purposes.

Definition 2: For $\mathsf{x} \in \mathcal{X}$ and $f\colon [0,1] \to \mathbb{R}$, define

$$I_f(\mathsf{x}) \triangleq \int f\left(\left|\tanh\left(\tfrac{\alpha}{2}\right)\right|\right) \mathsf{x}(d\alpha).$$

For $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}$, x_1 is said to be *degraded* with respect to x_2 (denoted $\mathsf{x}_1 \succeq \mathsf{x}_2$), if $I_f(\mathsf{x}_1) \geq I_f(\mathsf{x}_2)$ for all concave non-increasing f. Furthermore, x_1 is said to be *strictly degraded* with respect to x_2 (denoted $\mathsf{x}_1 \succ \mathsf{x}_2$) if $\mathsf{x}_1 \succeq \mathsf{x}_2$ and $\mathsf{x}_1 \neq \mathsf{x}_2$. We also write $\mathsf{x}_2 \preceq \mathsf{x}_1$ (respectively, $\mathsf{x}_2 \prec \mathsf{x}_1$) to mean $\mathsf{x}_1 \succeq \mathsf{x}_2$ (respectively, $\mathsf{x}_1 \succ \mathsf{x}_2$).

Recall that two measures x_1, x_2 are equal if $\mathsf{x}_1(E) = \mathsf{x}_2(E)$ for all Borel sets $E \subseteq \overline{\mathbb{R}}$. The class of concave non-increasing functions is rich enough to capture the notion of non-equality. That is, if $\mathsf{x}_1 \neq \mathsf{x}_2$, then there exists a concave non-increasing $f\colon [0,1] \to \mathbb{R}$ such that $I_f(\mathsf{x}_1) \neq I_f(\mathsf{x}_2)$.

Degradation defines a partial order on the space of symmetric probability measures, with the greatest element Δ_0 and the least element Δ_∞. Thus

$$\mathsf{x} \succ \Delta_\infty \text{ if } \mathsf{x} \neq \Delta_\infty, \text{ and } \mathsf{x} \prec \Delta_0 \text{ if } \mathsf{x} \neq \Delta_0.$$

This partial ordering is also preserved under the binary operations as follows.

Proposition 3: Suppose $\mathsf{x}_1, \mathsf{x}_2, \mathsf{x}_3 \in \mathcal{X}$.

i) If $\mathsf{x}_1 \succeq \mathsf{x}_2$, then

$$\mathsf{x}_1 * \mathsf{x}_3 \succeq \mathsf{x}_2 * \mathsf{x}_3, \quad \text{for all } \mathsf{x}_3 \in \mathcal{X}.$$

ii) The operators $\circledast$ and $\boxtimes$ also preserve a strict ordering for non-extremal measures. That is, if $\mathsf{x}_1 \succ \mathsf{x}_2$, then

$$\mathsf{x}_1 \circledast \mathsf{x}_3 \succ \mathsf{x}_2 \circledast \mathsf{x}_3 \quad \text{for } \mathsf{x}_3 \neq \Delta_\infty, \qquad \mathsf{x}_1 \boxtimes \mathsf{x}_3 \succ \mathsf{x}_2 \boxtimes \mathsf{x}_3 \quad \text{for } \mathsf{x}_3 \neq \Delta_0.$$

Proof. i) Direct application of [38, Lemma 4.80].

ii) It suffices to show that $\mathsf{x}_1 * \mathsf{x}_3 \neq \mathsf{x}_2 * \mathsf{x}_3$ under the stated conditions. For this, it is sufficient to construct a functional which gives different values under $\mathsf{x}_1 * \mathsf{x}_3$ and $\mathsf{x}_2 * \mathsf{x}_3$. The entropy functional (e.g., see Proposition 8(iv)) provides such a property.

□

Order by degradation is also preserved, much like the standard order of real numbers, under nonnegative multiplications and additions, i.e. for $0 \leq \alpha \leq 1$ and $\mathsf{x}_1 \succeq \mathsf{x}_2$, $\mathsf{x}_3 \succeq \mathsf{x}_4$,

$$\alpha \mathsf{x}_1 + (1-\alpha)\mathsf{x}_3 \succeq \alpha \mathsf{x}_2 + (1-\alpha)\mathsf{x}_4.$$

This ordering is our primary tool in describing relative channel quality. For further information see [38, pp. 204-208].

II.B.3 Entropy Functional for Symmetric Measures

To explicitly quantify the difference between two symmetric measures, one can employ the entropy functional. The entropy functional is the linear functional $\mathrm{H}\colon \mathcal{M} \to \mathbb{R}$ defined by

$$\mathrm{H}(\mathsf{x}) \triangleq \int \log_2\left(1 + e^{-\alpha}\right) \mathsf{x}(d\alpha).$$

This is the primary functional used in our analysis. It preserves the partial order under degradation and for $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}$, we have

$$\mathrm{H}(\mathsf{x}_1) > \mathrm{H}(\mathsf{x}_2) \quad \text{for } \mathsf{x}_1 \succ \mathsf{x}_2.$$

The restriction to symmetric probability measures also implies the bound

$$0 \leq \mathrm{H}(\mathsf{x}) \leq 1, \qquad \text{if } \mathsf{x} \in \mathcal{X}.$$

The operators $\circledast$ and $\circledbslash$ admit a number of relationships under the entropy functional. The following results will prove invaluable in the ensuing analysis. Proposition 4 provides an important conservation result (also known as the duality rule for entropy) and Proposition 5 extends this relation to encompass differences of symmetric probability measures.

Proposition 4 ([38, Lemma 4.41])**:** For $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}$,

$$\mathrm{H}\left(\mathsf{x}_1 \circledast \mathsf{x}_2\right) + \mathrm{H}\left(\mathsf{x}_1 \circledbslash \mathsf{x}_2\right) = \mathrm{H}\left(\mathsf{x}_1\right) + \mathrm{H}\left(\mathsf{x}_2\right).$$

Proposition 5: For x_1, x_2, x_3, $\mathsf{x}_4 \in \mathcal{X}$,

$$\begin{aligned}\mathrm{H}\left(\mathsf{x}_1 \circledast (\mathsf{x}_3 - \mathsf{x}_4)\right) + \mathrm{H}\left(\mathsf{x}_1 \circledbslash (\mathsf{x}_3 - \mathsf{x}_4)\right) &= \mathrm{H}\left(\mathsf{x}_3 - \mathsf{x}_4\right), \\ \mathrm{H}\left((\mathsf{x}_1 - \mathsf{x}_2) \circledast (\mathsf{x}_3 - \mathsf{x}_4)\right) + \mathrm{H}\left((\mathsf{x}_1 - \mathsf{x}_2) \circledbslash (\mathsf{x}_3 - \mathsf{x}_4)\right) &= 0.\end{aligned}$$

Proof. Consider the LHS of the first equality,

$$\begin{aligned}&\mathrm{H}\left(\mathsf{x}_1 \circledast (\mathsf{x}_3 - \mathsf{x}_4)\right) + \mathrm{H}\left(\mathsf{x}_1 \circledbslash (\mathsf{x}_3 - \mathsf{x}_4)\right) \\ &= \mathrm{H}\left(\mathsf{x}_1 \circledast \mathsf{x}_3\right) + \mathrm{H}\left(\mathsf{x}_1 \circledbslash \mathsf{x}_3\right) - \mathrm{H}\left(\mathsf{x}_1 \circledast \mathsf{x}_4\right) - \mathrm{H}\left(\mathsf{x}_1 \circledbslash \mathsf{x}_4\right) \\ &= \mathrm{H}\left(\mathsf{x}_1\right) + \mathrm{H}\left(\mathsf{x}_3\right) - \mathrm{H}\left(\mathsf{x}_1\right) - \mathrm{H}\left(\mathsf{x}_4\right) \quad \text{(Proposition 4)} \\ &= \mathrm{H}\left(\mathsf{x}_3 - \mathsf{x}_4\right).\end{aligned}$$

The second equality follows by expanding the LHS and applying the first equality twice. □

For $k \in \mathbb{N}$, let $M_k \colon \mathcal{M} \to \mathbb{R}$ denote the linear functional that maps $\mathsf{x} \in \mathcal{M}$ to its $2k$-th moment under tanh,

$$M_k(\mathsf{x}) \triangleq \int \tanh^{2k}\left(\tfrac{\alpha}{2}\right) \mathsf{x}(d\alpha).$$

Proposition 6: The following results hold.

i) For $\mathsf{x} \in \mathcal{X}$, $0 \leq M_k(\mathsf{x}) \leq 1$.

ii) For $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}$ with $\mathsf{x}_1 \succeq \mathsf{x}_2$, $M_k(\mathsf{x}_1) \leq M_k(\mathsf{x}_2)$.

iii) M_k satisfies the following product form identity for the operator $\boxtimes$,

$$M_k(\mathsf{x}_1 \boxtimes \mathsf{x}_2) = M_k(\mathsf{x}_1)M_k(\mathsf{x}_2).$$

iv) If $\mathsf{x} = \Delta_\infty$ (respectively, $\mathsf{x} = \Delta_0$), $M_k(\mathsf{x}) = 1$ (respectively, $M_k(\mathsf{x}) = 0$) for all k. Conversely, for some $\mathsf{x} \in \mathcal{X}$, if $M_k(\mathsf{x}) = 1$ (respectively, $M_k(\mathsf{x}) = 0$) for some k, then $\mathsf{x} = \Delta_\infty$ (respectively, $\mathsf{x} = \Delta_0$).

Proof. See Section II.H.2.2. □

Due to the symmetry of the measures, the entropy functional has an equivalent series representation in terms of the moments M_k.

Proposition 7 ([39, Lemma 3])**:** If $\mathsf{x} \in \mathcal{M}$, then

$$\mathrm{H}(\mathsf{x}) = \mathsf{x}\left(\overline{\mathbb{R}}\right) - \sum_{k=1}^{\infty} \gamma_k M_k(\mathsf{x}), \quad \text{where } \gamma_k = \frac{(\log 2)^{-1}}{2k(2k-1)}.$$

Proof. The main idea is to observe that

$$\log_2(1 + e^{-\alpha}) = 1 - \log_2(1 + \tanh(\tfrac{\alpha}{2})).$$

From there, use the series expansion of $\log_2(1+t)$ and Proposition 1 to combine the odd and even tanh moments. For a detailed proof, see [39, Lemma 3] and [38, pp. 267-268]. □

Proposition 8: From the series expansion for symmetric measures, the entropy functional satisfies the following properties.

i) For $\mathsf{y}_1, \mathsf{y}_2 \in \mathcal{X}_\mathrm{d}$,

$$\mathrm{H}(\mathsf{y}_1) = -\sum_{k=1}^{\infty} \gamma_k M_k(\mathsf{y}_1),$$
$$\mathrm{H}(\mathsf{y}_1 \boxtimes \mathsf{y}_2) = -\sum_{k=1}^{\infty} \gamma_k M_k(\mathsf{y}_1)M_k(\mathsf{y}_2).$$

ii) For $\mathsf{y} \in \mathcal{X}_{\mathrm{d}}$,

$$\mathrm{H}\left(\mathsf{y} \circledast \mathsf{y}\right) = -\sum_{k=1}^{\infty} \gamma_k M_k(\mathsf{y})^2 \leq 0, \qquad \mathrm{H}\left(\mathsf{y} \circledcirc \mathsf{y}\right) \geq 0.$$

with equality iff $\mathsf{y} = 0$. Additionally if $\mathsf{x} \in \mathcal{X}$,

$$\mathrm{H}\left(\mathsf{y} \circledast \mathsf{y} \circledast \mathsf{x}\right) \leq 0,$$

with equality iff $\mathsf{y} = 0$ or $\mathsf{x} = \Delta_0$.

iii) If $\mathsf{y}_1 = \mathsf{x}_1' - \mathsf{x}_1$, $\mathsf{y}_2 = \mathsf{x}_2' - \mathsf{x}_2$ with $\mathsf{x}_1' \succeq \mathsf{x}_1$, $\mathsf{x}_2' \succeq \mathsf{x}_2$,

$$\mathrm{H}\left(\mathsf{y}_1 \circledast \mathsf{y}_2\right) \leq 0, \qquad \mathrm{H}\left(\mathsf{y}_1 \circledcirc \mathsf{y}_2\right) \geq 0.$$

iv) If $\mathsf{x}_1 \succ \mathsf{x}_2$, then

$$\begin{aligned} \mathrm{H}\left(\mathsf{x}_1 \circledcirc \mathsf{x}_3\right) &> \mathrm{H}\left(\mathsf{x}_2 \circledcirc \mathsf{x}_3\right) \quad \text{if } \mathsf{x}_3 \neq \Delta_\infty \\ \mathrm{H}\left(\mathsf{x}_1 \circledast \mathsf{x}_3\right) &> \mathrm{H}\left(\mathsf{x}_2 \circledast \mathsf{x}_3\right) \quad \text{if } \mathsf{x}_3 \neq \Delta_0. \end{aligned}$$

Proof. See Section II.H.2.3. □

Proposition 8 also implies the following upper bound on the entropy functional for differences of symmetric probability measures under the operators $\circledcirc$ and $\circledast$.

Proposition 9: For $\mathsf{x}_1, \mathsf{x}_1', \mathsf{x}_2, \mathsf{x}_3, \mathsf{x}_4 \in \mathcal{X}$ with $\mathsf{x}_1' \succeq \mathsf{x}_1$,

$$\begin{aligned} \left|\mathrm{H}\left((\mathsf{x}_1' - \mathsf{x}_1) * (\mathsf{x}_2 - \mathsf{x}_3)\right)\right| &\leq \mathrm{H}\left(\mathsf{x}_1' - \mathsf{x}_1\right), \\ \left|\mathrm{H}\left((\mathsf{x}_1' - \mathsf{x}_1) * (\mathsf{x}_2 - \mathsf{x}_3) * \mathsf{x}_4\right)\right| &\leq \mathrm{H}\left(\mathsf{x}_1' - \mathsf{x}_1\right). \end{aligned}$$

Proof. Consider the first inequality with the operator $\circledast$. From Proposition 8(i),

$$\begin{aligned} \left|\mathrm{H}\left((\mathsf{x}_1' - \mathsf{x}_1) \circledast (\mathsf{x}_2 - \mathsf{x}_3)\right)\right| &\leq \sum_{k=1}^{\infty} \gamma_k \left|M_k(\mathsf{x}_1' - \mathsf{x}_1)\right| \left|M_k(\mathsf{x}_2 - \mathsf{x}_3)\right| \\ &\overset{(a)}{=} -\sum_{k=1}^{\infty} \gamma_k M_k(\mathsf{x}_1' - \mathsf{x}_1) \left|M_k(\mathsf{x}_2 - \mathsf{x}_3)\right| \end{aligned}$$

$$\overset{(b)}{\leq} -\sum_{k=1}^{\infty} \gamma_k M_k(\mathsf{x}_1' - \mathsf{x}_1)$$
$$= \mathrm{H}\left(\mathsf{x}_1' - \mathsf{x}_1\right),$$

where (a) follows from $M_k(\mathsf{x}_1') \leq M_k(\mathsf{x}_1)$ and (b) follows since $0 \leq M_k(\mathsf{x}_2), M_k(\mathsf{x}_3) \leq 1$. The result for the operator $\circledast$ then follows from Proposition 5. The second inequality follows from the first by replacing x_2, x_3 with $\mathsf{x}_2 * \mathsf{x}_4$, $\mathsf{x}_3 * \mathsf{x}_4$. □

The series expansion in Proposition 7 leads us to define the following metric on the set of symmetric probability measures.

Definition 10: For $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}$, the *entropy distance* is defined as

$$d_{\mathrm{H}}(\mathsf{x}_1, \mathsf{x}_2) = \sum_{k=1}^{\infty} \gamma_k \left| M_k(\mathsf{x}_1) - M_k(\mathsf{x}_2) \right|.$$

When $\mathsf{x}_2 \succeq \mathsf{x}_1$, observe that $d_{\mathrm{H}}(\mathsf{x}_1, \mathsf{x}_2) = \mathrm{H}\left(\mathsf{x}_2 - \mathsf{x}_1\right)$; hence the name entropy distance. Thus, $d_{\mathrm{H}}(\Delta_\infty, \mathsf{x}) = \mathrm{H}\left(\mathsf{x}\right)$ and $d_{\mathrm{H}}(\mathsf{x}, \Delta_0) = 1 - \mathrm{H}\left(\mathsf{x}\right)$. Moreover, for any $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}$, $d_{\mathrm{H}}(\mathsf{x}_1, \mathsf{x}_2) \geq \left|\mathrm{H}\left(\mathsf{x}_1 - \mathsf{x}_2\right)\right|$, and for $\mathsf{x}_3 \succeq \mathsf{x}_2 \succeq \mathsf{x}_1$, $d_{\mathrm{H}}(\mathsf{x}_1, \mathsf{x}_3) \geq d_{\mathrm{H}}(\mathsf{x}_1, \mathsf{x}_2)$.

Proposition 11: We have the following topological results related to the entropy distance.

i) The entropy distance d_{H} is a metric on the set of symmetric probability measures, $\mathcal{X}$.

ii) The metric topology $(\mathcal{X}, d_{\mathrm{H}})$ is compact.

iii) The entropy functional $\mathrm{H}\colon \mathcal{X} \to [0, 1]$ is continuous.

iv) With the product topology on $\mathcal{X} \times \mathcal{X}$, the operators $\circledast\colon \mathcal{X} \times \mathcal{X} \to \mathcal{X}$ and $\circledcirc\colon \mathcal{X} \times \mathcal{X} \to \mathcal{X}$ are continuous.

v) If a sequence of measures $\{\mathsf{x}_n\}_{n=1}^{\infty}$ in $\mathcal{X}$ satisfies $\mathsf{x}_n \succeq \mathsf{x}_{n-1}$ (respectively, $\mathsf{x}_n \preceq \mathsf{x}_{n-1}$), then $\mathsf{x}_n \xrightarrow{d_{\mathrm{H}}} \mathsf{x}$, for some $\mathsf{x} \in \mathcal{X}$, and $\mathsf{x} \succeq \mathsf{x}_n$ (respectively, $\mathsf{x} \preceq \mathsf{x}_n$) for all n.

vi) If $\mathsf{x}_n' \succeq \mathsf{x}_n$ and $\mathsf{x}_n' \xrightarrow{d_{\mathrm{H}}} \mathsf{x}'$, $\mathsf{x}_n \xrightarrow{d_{\mathrm{H}}} \mathsf{x}$, then $\mathsf{x}' \succeq \mathsf{x}$.

Proof. See Section II.H.1. □

We use these topological results minimally. The compactness of $\mathcal{X}$ and the continuity of $\mathrm{H}(\cdot)$, $\circledast$ and $\boxledast$ are used to establish the existence of minimizing measures for some functionals. These minima are used to show the threshold saturation converse for LDPC ensembles. For the achievability result (Theorems 44 and 61), we require properties (v) and (vi) in the above proposition, which appear in [38, Section 4.1]. We note that our previous article, [40], shows the achievability of threshold saturation for LDPC ensembles using only existing convergence results from [38, Section 4.1].

II.B.4 Bhattacharyya Functional for Symmetric Measures

The quantity that characterizes the stability of LDPC ensembles is the Bhattacharyya functional, $\mathfrak{B}\colon \mathcal{M} \to \mathbb{R}$,

$$\mathfrak{B}(\mathsf{x}) \triangleq \int e^{-\alpha/2} \mathsf{x}(d\alpha).$$

Since this is a Laplace transform of the measure evaluated at $1/2$, Bhattacharyya functional is multiplicative under the convolution operator $\circledast$,

$$\mathfrak{B}(\mathsf{x}^{\circledast n}) = \mathfrak{B}(\mathsf{x})^n.$$

Like the entropy functional, the Bhattacharyya functional also preserves the degradation order,

$$\mathfrak{B}(\mathsf{x}_1) > \mathfrak{B}(\mathsf{x}_2), \quad \text{if } \mathsf{x}_1 \succ \mathsf{x}_2.$$

It also satisfies the bound

$$0 \leq \mathfrak{B}(\mathsf{x}) \leq 1, \quad \text{if } \mathsf{x} \in \mathcal{X}.$$

Importantly, the Bhattacharyya functional characterizes the logarithmic decay rate of the entropy functional under the operator $\circledast$.

Proposition 12: For $\mathsf{x} \in \mathcal{X}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{H}\left(\mathsf{x}^{\circledast n}\right) = \log \mathfrak{B}(\mathsf{x}).$$

Proof. See Section II.H.2.4. □

II.B.5 Directional Derivatives

The main result in this chapter is derived using potential theory and differential relations. One can avoid some technical challenges of differentiation in the abstract space of measures by focusing on directional derivatives of functionals that map measures to real numbers.

Definition 13: Let $F\colon \mathcal{M} \to \mathbb{R}$ be a functional on $\mathcal{M}$. The *directional derivative* of F at x in the direction y is

$$\mathrm{d}_{\mathsf{x}} F(\mathsf{x})[\mathsf{y}] \triangleq \lim_{\delta \to 0} \frac{F(\mathsf{x} + \delta \mathsf{y}) - F(\mathsf{x})}{\delta},$$

whenever the limit exists. For $G\colon \mathcal{M} \to \mathcal{M}$, define

$$\mathrm{d}_{\mathsf{x}} F(G(\mathsf{x}))[\mathsf{y}] \triangleq \mathrm{d}_{\mathsf{x}} (F \circ G)(\mathsf{x})[\mathsf{y}] = \lim_{\delta \to 0} \frac{F(G(\mathsf{x} + \delta \mathsf{y})) - F(G(\mathsf{x}))}{\delta},$$

whenever the limit exists. For convenience, we sometimes write

$$\mathrm{d}_{\mathsf{x}} F(\mathsf{x})[\mathsf{y}] \Big|_{\mathsf{x}=\mathsf{x}_1} \triangleq \mathrm{d}_{\mathsf{x}_1} F(\mathsf{x}_1)[\mathsf{y}].$$

This definition is naturally extended to higher-order directional derivatives using

$$\mathrm{d}^n_{\mathsf{x}} F(\mathsf{x})[\mathsf{y}_1, \ldots, \mathsf{y}_n] \triangleq \mathrm{d}_{\mathsf{x}} \left(\cdots \mathrm{d}_{\mathsf{x}} \left(\mathrm{d}_{\mathsf{x}} F(\mathsf{x}) [\mathsf{y}_1] \right) [\mathsf{y}_2] \cdots \right) [\mathsf{y}_n],$$

and vectors of measures using, for $\underline{\mathsf{x}} = [\mathsf{x}_1, \cdots, \mathsf{x}_m]$,

$$\mathrm{d}_{\underline{\mathsf{x}}} F(\underline{\mathsf{x}})[\underline{\mathsf{y}}] \triangleq \lim_{\delta \to 0} \frac{F(\underline{\mathsf{x}} + \delta \underline{\mathsf{y}}) - F(\underline{\mathsf{x}})}{\delta},$$

whenever the limit exists. Similarly, we can define higher-order directional derivatives for the composition of functions and functionals on vectors of measures.

The utility of directional derivatives for linear functionals is evident from the following result.

Proposition 14: Let $F\colon \mathcal{M} \to \mathbb{R}$ be a linear functional, and $*$ be either $\circledast$ or $\boxplus$.

Then, for $\mathsf{x}, \mathsf{y}, \mathsf{z} \in \mathcal{M}$, we have

$$\begin{aligned} \mathrm{d}_{\mathsf{x}} F(\mathsf{x}^{*n})[\mathsf{y}] &= nF(\mathsf{x}^{*(n-1)} * \mathsf{y}), \\ \mathrm{d}^2_{\mathsf{x}} F(\mathsf{x}^{*n})[\mathsf{y}, \mathsf{z}] &= n\,(n-1)\,F\left(\mathsf{x}^{*(n-2)} * \mathsf{y} * \mathsf{z}\right). \end{aligned}$$

Proof. Associativity, commutativity, and linearity of the binary operator $*$ allow a binomial expansion of $(\mathsf{x} + \delta\mathsf{y})^{*n}$:

$$(\mathsf{x} + \delta\mathsf{y})^{*n} = \sum_{i=0}^{n} \delta^i \binom{n}{i} \mathsf{x}^{*(n-i)} * \mathsf{y}^{*i}.$$

Then, the linearity of F implies that

$$F\left((\mathsf{x} + \delta\mathsf{y})^{*n}\right) - F\left(\mathsf{x}^{*n}\right) = \delta n F(\mathsf{x}^{*(n-1)} * \mathsf{y}) + \sum_{i=2}^{n} \delta^i \binom{n}{i} F(\mathsf{x}^{*(n-i)} * \mathsf{y}^{*i}).$$

Dividing by δ and taking a limit gives

$$\mathrm{d}_{\mathsf{x}} F(\mathsf{x}^{*n})[\mathsf{y}] = nF(\mathsf{x}^{*(n-1)} * \mathsf{y}).$$

An analogous argument shows that

$$\mathrm{d}^2_{\mathsf{x}} F(\mathsf{x}^{*n})[\mathsf{y}, \mathsf{z}] = n(n-1)F(\mathsf{x}^{*(n-2)} * \mathsf{y} * \mathsf{z}).$$

□

In the following proposition, we evaluate the directional derivative of a linear functional which contains both the operators $\circledast$ and $\boxtimes$.

Proposition 15: Suppose $F\colon \mathcal{M} \to \mathbb{R}$ is a linear functional and p, q are polynomials. Then

$$\mathrm{d}_{\mathsf{x}} F(p^{\circledast}(q^{\boxtimes}(\mathsf{x})))[\mathsf{y}] = F\left(p'^{\circledast}\left(q^{\boxtimes}(\mathsf{x})\right) \circledast \left(q'^{\boxtimes}(\mathsf{x}) \boxtimes \mathsf{y}\right)\right).$$

Proof. Since F is a linear functional, it suffices to show the result when $p(\alpha) = \alpha^n$. In view of the proof of previous proposition, the coefficient of δ in

$$(q^{\boxtimes}(\mathsf{x} + \delta\mathsf{y}))^{\circledast n} - (q^{\boxtimes}(\mathsf{x}))^{\circledast n}$$

determines the first-order directional derivative. Again, from the binomial expansion,

$$\begin{aligned}(q^{\boxtimes}(\mathsf{x}+\delta\mathsf{y}))^{\circledast n} - (q^{\boxtimes}(\mathsf{x}))^{\circledast n} &= \Big(\sum_{k=0}^{\deg(\mathrm{q})} q_k(\mathsf{x}+\delta\mathsf{y})^{\boxtimes k}\Big)^{\circledast n} - (q^{\boxtimes}(\mathsf{x}))^{\circledast n} \\ &= \Big(q^{\boxtimes}(\mathsf{x}) + \Big(\sum_{k=1}^{\deg(q)} k q_k \mathsf{x}^{\boxtimes k-1} \circledast \mathsf{y}\Big)\delta + o(\delta)\Big)^{\circledast n} - (q^{\boxtimes}(\mathsf{x}))^{\circledast n} \\ &= \left(q^{\boxtimes}(\mathsf{x}) + \left(q'^{\boxtimes}(\mathsf{x}) \circledast \mathsf{y}\right)\delta + o(\delta)\right)^{\circledast n} - (q^{\boxtimes}(\mathsf{x}))^{\circledast n}\end{aligned}$$

A direct inspection from the multinomial expansion of the first term gives the coefficient of δ,

$$n\left(\left(q^{\boxtimes}(\mathsf{x})\right)^{\circledast n-1}\right) \circledast (q'^{\boxtimes}(\mathsf{x}) \circledast \mathsf{y}).$$

Thus, when $p(\alpha) = \alpha^n$,

$$\mathrm{d}_{\mathsf{x}} F(p^{\circledast}(q^{\boxtimes}(\mathsf{x})))[\mathsf{y}] = F\left(p'^{\circledast}\left(q^{\boxtimes}(\mathsf{x})\right) \circledast \left(q'^{\boxtimes}(\mathsf{x}) \circledast \mathsf{y}\right)\right).$$

The general result follows. □

One recurring theme here is when relating two quantities $F(\mathsf{x}_1)$, $F(\mathsf{x}_2)$ is to consider a parameterized path from x_1 to x_2, of the form $\mathsf{x}_1 + t(\mathsf{x}_2 - \mathsf{x}_1) = (1-t)\mathsf{x}_1 + t\mathsf{x}_2$, in the set of symmetric probability measures, and analyze the directional derivative of $F(\cdot)$ at $\mathsf{x}_1 + t(\mathsf{x}_2 - \mathsf{x}_1)$, in the direction $\mathsf{x}_2 - \mathsf{x}_1$. The following proposition formalizes this idea.

Proposition 16: Let $F\colon \mathcal{X} \to \mathbb{R}$ be a linear functional, $*$ either $\circledast$ or $\boxtimes$, p a polynomial, and $G\colon \mathcal{X} \to \mathbb{R}$, $G(\mathsf{x}) = F(p^*(\mathsf{x}))$. For $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}$, let $\phi\colon [0,1] \to \mathbb{R}$,

$$\phi(t) = G(\mathsf{x}_1 + t(\mathsf{x}_2 - \mathsf{x}_1)).$$

Then, $\phi(t)$ is a polynomial in t,

$$\begin{aligned}\phi'(t) &= \mathrm{d}_{\mathsf{x}} G(\mathsf{x})[\mathsf{x}_2 - \mathsf{x}_1]\Big|_{\mathsf{x}=\mathsf{x}_1+t(\mathsf{x}_2-\mathsf{x}_1)}, \text{ and} \\ \phi''(t) &= \mathrm{d}^2_{\mathsf{x}} G(\mathsf{x})[\mathsf{x}_2 - \mathsf{x}_1, \mathsf{x}_2 - \mathsf{x}_1]\Big|_{\mathsf{x}=\mathsf{x}_1+t(\mathsf{x}_2-\mathsf{x}_1)}.\end{aligned}$$

Proof. Since $\mathsf{x}_1 + t(\mathsf{x}_2 - \mathsf{x}_1) = (1-t)\mathsf{x}_1 + t\mathsf{x}_2$, from the binomial expansion,

$$(\mathsf{x}_1 + t(\mathsf{x}_2 - \mathsf{x}_1))^{*n} = \sum_{k=0}^{n} \binom{n}{k} \left(\mathsf{x}_1^{*n-k} * \mathsf{x}_2^{*k}\right) (1-t)^{n-k} t^k.$$

Since F is a linear functional,

$$\begin{aligned}
\phi(t) &= G(\mathsf{x}_1 + t(\mathsf{x}_2 - \mathsf{x}_1)) \\
&= F(p^*(\mathsf{x}_1 + t(\mathsf{x}_2 - \mathsf{x}_1))) \\
&= \sum_{n=0}^{\deg(p)} p_n F\left((\mathsf{x}_1 + t(\mathsf{x}_2 - \mathsf{x}_1)^{*n}\right) \\
&= \sum_{n=0}^{\deg(p)} p_n \sum_{k=0}^{n} \binom{n}{k} F\left(\mathsf{x}_1^{*n-k} * \mathsf{x}_2^{*k}\right) (1-t)^{n-k} t^k,
\end{aligned}$$

is a polynomial of degree at most $\deg(p)$. Moreover,

$$\begin{aligned}
\phi'(t) &= \lim_{\delta \to 0} \frac{G(\mathsf{x}_1 + (t+\delta)(\mathsf{x}_2 - \mathsf{x}_1)) - G(\mathsf{x}_1 + t(\mathsf{x}_2 - \mathsf{x}_1))}{\delta} \\
&= \mathrm{d}_{\mathsf{x}} G(\mathsf{x})[\mathsf{x}_2 - \mathsf{x}_1] \Big|_{\mathsf{x} = \mathsf{x}_1 + t(\mathsf{x}_2 - \mathsf{x}_1)},
\end{aligned}$$

by Definition 13. The expression for second derivative $\phi''(t)$ follows similarly. □

As such, if $\phi'(t) \leq 0$ in the above proposition for all $t \in (0,1)$, we find that $G(\mathsf{x}_1) \leq G(\mathsf{x}_2)$ because $\phi(0) = G(\mathsf{x}_1)$, $\phi(1) = G(\mathsf{x}_2)$.

Remark 17: In general, applying Taylor's theorem to some mapping $F\colon \mathcal{X} \to \mathcal{X}$ requires Fréchet derivatives. However, the linearity of the entropy functional and its interplay with the operators $\circledast$ and $\boxplus$ impose a polynomial structure on the functions of interest, obviating the need for advanced mathematical machinery. Therefore, Taylor's theorem becomes quite simple for parameterized linear functionals $\phi\colon [0,1] \to \mathbb{R}$ of the form

$$\phi(t) = F\left(\mathsf{x}_1 + t(\mathsf{x}_2 - \mathsf{x}_1)\right).$$

II.C LOW-DENSITY PARITY-CHECK ENSEMBLES

II.C.1 Single System

Let LDPC(λ, ρ) denote the LDPC ensemble with variable-node degree distribution λ and check-node degree distribution ρ. The edge perspective degree distributions λ, ρ have an equivalent representation in terms of the node perspective degree distributions L, R given by

$$\lambda(t) = \frac{L'(t)}{L'(1)}, \qquad\qquad \rho(t) = \frac{R'(t)}{R'(1)}.$$

It is important to note that the distributions λ, ρ, L and R are all *polynomials*. We assume that the LDPC(λ, ρ) ensemble does not have any degree-one variable-nodes, as these ensembles exhibit non-negligible error floors. We also refer to this ensemble as a single system to differentiate from its coupled variant introduced later.

Density evolution (DE) characterizes the asymptotic performance of the LDPC (λ, ρ) ensemble under message-passing decoding by describing the evolution of message distributions with iteration. Under locally optimal processing, the message-passing decoder is equivalent to the belief-propagation (BP) decoder. For the LDPC (λ, ρ) ensemble, the DE under BP decoding is described by

$$\tilde{\mathsf{x}}^{(\ell+1)} = \mathsf{c} \circledast \lambda^{\circledast}(\rho^{\boxplus}(\tilde{\mathsf{x}}^{(\ell)})), \tag{II.1}$$

where $\tilde{\mathsf{x}}^{(\ell)}$ is the variable-node output distribution after ℓ iterations of message passing [38, 41]. If the iterative system in (II.1) is initialized with $\mathsf{x}^{(0)} = \mathsf{a}$, the variable-node output-distribution after ℓ iterations of message-passing is denoted by $\mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a}; \mathsf{c})$. The variable-node output after one iteration is also denoted by

$$\mathsf{T}_{\mathrm{s}}(\mathsf{a}; \mathsf{c}) \triangleq \mathsf{T}_{\mathrm{s}}^{(1)}(\mathsf{a}; \mathsf{c}) = \mathsf{c} \circledast \lambda^{\circledast}(\rho^{\boxplus}(\mathsf{a})).$$

If the sequence of measures $\{\mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a}; \mathsf{c})\}$ converges in $(\mathcal{X}, d_{\mathrm{H}})$, then its limit is denoted by $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a}; \mathsf{c})$.

The DE update operator T_{s} satisfies certain monotonicity properties. These properties play a crucial role in the analysis of LDPC ensembles.

Lemma 18 ([38, Section 4.6])**:** The operator $\mathsf{T}_{\mathrm{s}}^{(\ell)} \colon \mathcal{X} \times \mathcal{X} \to \mathcal{X}$ satisfies the following monotonicity properties for all $1 \leq \ell < \infty$.

i) If $\mathsf{a}_1 \succeq \mathsf{a}_2$, then $\mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a}_1;\mathsf{c}) \succeq \mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a}_2;\mathsf{c})$ for all $\mathsf{c} \in \mathcal{X}$.

ii) If $\mathsf{c}_1 \succeq \mathsf{c}_2$, $\mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a};\mathsf{c}_1) \succeq \mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a};\mathsf{c}_2)$ for all $\mathsf{a} \in \mathcal{X}$.

iii) If $\mathsf{T}_{\mathrm{s}}(\mathsf{a};\mathsf{c}) \preceq \mathsf{a}$, then $\mathsf{T}_{\mathrm{s}}^{(\ell+1)}(\mathsf{a};\mathsf{c}) \preceq \mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a};\mathsf{c})$. Moreover, $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c})$ exists and satisfies $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c}) \preceq \mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a};\mathsf{c})$,

$$\mathsf{T}_{\mathrm{s}}(\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c});\mathsf{c}) = \mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c}).$$

iv) If $\mathsf{T}_{\mathrm{s}}(\mathsf{a};\mathsf{c}) \succeq \mathsf{a}$, then $\mathsf{T}_{\mathrm{s}}^{(\ell+1)}(\mathsf{a};\mathsf{c}) \succeq \mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a};\mathsf{c})$. Moreover, $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c})$ exists and satisfies $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c}) \succeq \mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a};\mathsf{c})$,

$$\mathsf{T}_{\mathrm{s}}(\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c});\mathsf{c}) = \mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c}).$$

Proof. The monotonicity properties can be derived from Proposition 3, while the existence of the limit in $(\mathcal{X}, d_{\mathrm{H}})$ and its properties follow from Proposition 11. That the limit satisfies

$$\mathsf{T}_{\mathrm{s}}(\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c});\mathsf{c}) = \mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c})$$

follows from the continuity of $\circledast$, $\boxtimes$, and the fact that λ, ρ are polynomials. □

Thus, when (II.1) is initialized with Δ_0, the sequence of measures $\{\mathsf{T}_{\mathrm{s}}^{(\ell)}(\Delta_0;\mathsf{c})\}$, satisfies $\mathsf{T}_{\mathrm{s}}(\Delta_0;\mathsf{c}) \preceq \Delta_0$, and converges to a limit x, which satisfies

$$\mathsf{x} = \mathsf{c} \circledast \lambda^{\circledast}(\rho^{\boxtimes}(\mathsf{x})).$$

Definition 19: A measure $\mathsf{x} \in \mathcal{X}$ is a DE *fixed point* for the LDPC(λ, ρ) ensemble if

$$\mathsf{x} = \mathsf{c} \circledast \lambda^{\circledast}\left(\rho^{\boxtimes}(\mathsf{x})\right).$$

We now state some necessary definitions for the single system potential framework. Included are the potential functional, stationary points, the directional derivative of the potential functional, and thresholds.

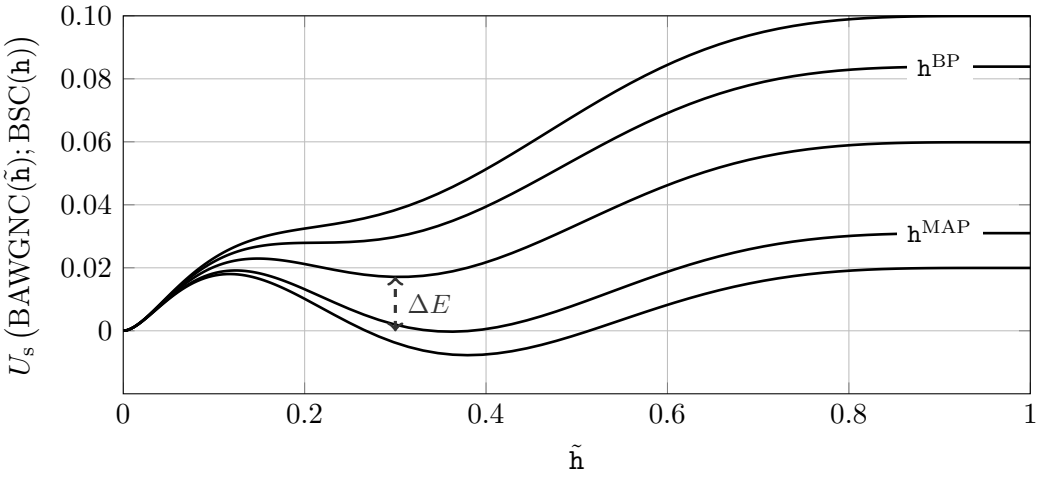

Figure II.1: Potential functional for the LDPC ensemble with $(\lambda(t), \rho(t)) = (t^2, t^5)$ over a BSC. The values of $\mathtt{h}$ for these curves are, from the top to bottom, 0.40, 0.416, 0.44, 0.469, 0.48. The other input to the potential functional is the LLR distribution for the binary AWGN channel (BAWGNC) with entropy $\tilde{\mathtt{h}}$. The choice of BAWGNC distribution is arbitrary.

Definition 20: The *potential functional*, $U_{\mathrm{s}} \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, of the LDPC$(\lambda, \rho)$ ensemble and a channel $\mathsf{c} \in \mathcal{X}$ is

$$\begin{aligned} U_{\mathrm{s}}(\mathsf{x}; \mathsf{c}) \triangleq \tfrac{L'(1)}{R'(1)} \mathrm{H}\left(R^{\boxledast}(\mathsf{x})\right) + L'(1)\mathrm{H}\left(\rho^{\boxledast}(\mathsf{x})\right) \\ - L'(1)\mathrm{H}\left(\mathsf{x} \boxledast \rho^{\boxledast}(\mathsf{x})\right) - \mathrm{H}\left(\mathsf{c} \circledast L^{\circledast}\left(\rho^{\boxledast}(\mathsf{x})\right)\right). \end{aligned}$$

Remark 21: The potential functional is essentially the negative of the trial entropy, formally known as the replica-symmetric free entropy, calculated in [36, 39, 42].[1] In Section II.H.7, we describe the Bethe formalism to obtain the free entropy and detail the calculations involved to derive the potential in Definition 20. When applied to the binary erasure channel, U_{s} is a constant multiple of the potential function defined in [11]. An example of $U_{\mathrm{s}}(\mathsf{x}; \mathsf{c})$ is shown in Fig. II.1.

It is hard to define precisely what conditions are required for a potential functional, that operates on measures, to prove threshold saturation. But, the crucial properties of the single system potential that we leverage are 1) the fixed points of

[1] While it is possible to use the term replica-symmetric free entropy instead of 'potential', our terminology is consistent with [11, 12, 35]. Moreover, we later define *coupled potential*; this brings both definitions together. In addition, for general systems, potential function need not be defined from the free entropy (e.g., see [28]).

the single system DE are the stationary points of the single system potential (Lemma 23), 2) there exists a spatially-coupled potential, defined by a spatial average of the single system potential (Definition 37), where the fixed points of spatially-coupled DE are stationary points of the spatially-coupled potential (Lemma 38).

The entropy functional and the operators ($\circledast$, $\boxtimes$) are continuous. Hence, the potential functional $U_{\mathrm{s}}(\cdot\,;\mathsf{c})$ for a fixed c is continuous. Since the metric topology $(\mathcal{X}, d_{\mathrm{H}})$ is compact, $U_{\mathrm{s}}(\cdot\,;\mathsf{c})$ achieves its minimum and maximum on $\mathcal{X}$. Though we also have the joint continuity of $U_{\mathrm{s}}(\cdot\,;\cdot)$, it is not used in this work.

Definition 22: A measure $\mathsf{x} \in \mathcal{X}$ is a *stationary point* of the potential if, for all $\mathsf{y} \in \mathcal{X}_{\mathrm{d}}$,

$$\mathrm{d}_{\mathsf{x}} U_{\mathrm{s}}(\mathsf{x};\mathsf{c})[\mathsf{y}] = 0.$$

Lemma 23: For x, $\mathsf{c} \in \mathcal{X}$ and $\mathsf{y} \in \mathcal{X}_{\mathrm{d}}$, the directional derivative of the potential functional with respect to x in the direction y is

$$\mathrm{d}_{\mathsf{x}} U_{\mathrm{s}}(\mathsf{x};\mathsf{c})[\mathsf{y}] = L'(1)\mathrm{H}\left(\left[\mathsf{T}_{\mathrm{s}}(\mathsf{x};\mathsf{c}) - \mathsf{x}\right] \boxtimes \left[\rho'^{\boxtimes}(\mathsf{x}) \boxtimes \mathsf{y}\right]\right).$$

Proof. Since the distributions λ, ρ, L, R are polynomials, the directional derivative for each of the four terms can be calculated following the procedure outlined in the proof of Proposition 14. The directional derivatives of the first three terms are

$$\begin{aligned}
\mathrm{d}_{\mathsf{x}} \mathrm{H}\left(R^{\boxtimes}(\mathsf{x})\right)[\mathsf{y}] &= R'(1)\mathrm{H}\left(\rho^{\boxtimes}(\mathsf{x}) \boxtimes \mathsf{y}\right), \\
\mathrm{d}_{\mathsf{x}} \mathrm{H}\left(\rho^{\boxtimes}(\mathsf{x})\right)[\mathsf{y}] &= \mathrm{H}\left(\rho'^{\boxtimes}(\mathsf{x}) \boxtimes \mathsf{y}\right), \\
\mathrm{d}_{\mathsf{x}} \mathrm{H}\left(\mathsf{x} \boxtimes \rho^{\boxtimes}(\mathsf{x})\right)[\mathsf{y}] &= \mathrm{H}\left(\rho^{\boxtimes}(\mathsf{x}) \boxtimes \mathsf{y}\right) + \mathrm{H}\left(\mathsf{x} \boxtimes \rho'^{\boxtimes}(\mathsf{x}) \boxtimes \mathsf{y}\right) \\
&\overset{(a)}{=} \mathrm{H}\left(\rho^{\boxtimes}(\mathsf{x}) \boxtimes \mathsf{y}\right) + \mathrm{H}\left(\rho'^{\boxtimes}(\mathsf{x}) \boxtimes \mathsf{y}\right) - \mathrm{H}\left(\mathsf{x} \circledast \left[\rho'^{\boxtimes}(\mathsf{x}) \boxtimes \mathsf{y}\right]\right),
\end{aligned}$$

where (a) follows from Proposition 5 with the observation that $\rho'^{\boxtimes}(\mathsf{x}) \boxtimes \mathsf{y}$ is a difference of probability measures multiplied by the scalar $\rho'(1)$. Since the operators $\circledast$ and $\boxtimes$ do not associate, one must exercise care in analyzing the last term. From Proposition 15,

$$\mathrm{d}_{\mathsf{x}} \mathrm{H}\left(\mathsf{c} \circledast L^{\circledast}\left(\rho^{\boxtimes}(\mathsf{x})\right)\right)[\mathsf{y}] = L'(1)\mathrm{H}\left(\left[\mathsf{c} \circledast \lambda^{\circledast}(\rho^{\boxtimes}(\mathsf{x}))\right] \circledast \left[\rho'^{\boxtimes}(\mathsf{x}) \boxtimes \mathsf{y}\right]\right).$$

Consolidating the four terms,

$$\mathrm{d}_{\mathsf{x}} U_{\mathrm{s}}(\mathsf{x};\mathsf{c})[\mathsf{y}] = L'(1)\mathrm{H}\left([\mathsf{x} - \mathsf{T}_{\mathrm{s}}(\mathsf{x};\mathsf{c})] \circledast \left[\rho'^{\boxplus}(\mathsf{x}) \boxplus \mathsf{y}\right]\right).$$

Using Proposition 5, we have the desired result. □

Lemma 24: If $\mathsf{x} \in \mathcal{X}$ is a fixed point of single system DE, then it is also a stationary point of the potential functional. Moreover, for a fixed channel c, the minimum of the potential functional,

$$\min_{\mathsf{x}\in\mathcal{X}} U_{\mathrm{s}}(\mathsf{x};\mathsf{c}),$$

occurs only at a fixed point of single system DE.

Proof. See Section II.H.3.1. □

Definition 25: For the LDPC(λ, ρ) ensemble and a channel $\mathsf{c} \in \mathcal{X}$, define

i) The *basin of attraction to* Δ_∞ as

$$\mathcal{V}(\mathsf{c}) \triangleq \left\{\mathsf{a} \in \mathcal{X} \mid \mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c}) = \Delta_\infty\right\}.$$

ii) The *energy gap* as

$$\Delta E(\mathsf{c}) \triangleq \inf_{\mathsf{x}\in\mathcal{X}\setminus\mathcal{V}(\mathsf{c})} U_{\mathrm{s}}(\mathsf{x};\mathsf{c}),$$

with the convention that the infimum over the empty set is ∞.

The only fixed point contained in $\mathcal{V}(\mathsf{c})$ is the trivial Δ_∞ fixed point. Therefore, all other fixed points are in the complement, $\mathcal{X} \setminus \mathcal{V}(\mathsf{c})$.

Lemma 26: Suppose $\mathsf{c}_1 \succ \mathsf{c}_2$. Then

i) $U_{\mathrm{s}}(\mathsf{x};\mathsf{c}_1) < U_{\mathrm{s}}(\mathsf{x};\mathsf{c}_2)$ if $\mathsf{x} \neq \Delta_\infty$

ii) $\mathcal{V}(\mathsf{c}_1) \subseteq \mathcal{V}(\mathsf{c}_2)$ and $\mathcal{X} \setminus \mathcal{V}(\mathsf{c}_1) \supseteq \mathcal{X} \setminus \mathcal{V}(\mathsf{c}_2)$

iii) $\Delta E(\mathsf{c}_1) \leq \Delta E(\mathsf{c}_2)$

Proof. See Section II.H.3.2. □

Definition 27: A *family of BMS channels* is a function $\mathtt{c}(\cdot)\colon [0,1] \to \mathcal{X}$ that is

i) ordered by degradation, $\mathtt{c}(\mathtt{h}_1) \succeq \mathtt{c}(\mathtt{h}_2)$ for $\mathtt{h}_1 \geq \mathtt{h}_2$,

ii) parameterized by entropy $\mathrm{H}\left(\mathtt{c}(\mathtt{h})\right) = \mathtt{h}$.

Definition 28: Consider a family of BMS channels and the LDPC(λ, ρ) ensemble. Define

i) The *BP threshold* as

$$\mathtt{h}^{\mathrm{BP}} \triangleq \sup\left\{\mathtt{h} \in [0,1] \mid \mathsf{T}_{\mathrm{s}}^{(\infty)}(\Delta_0; \mathtt{c}(\mathtt{h})) = \Delta_\infty\right\}.$$

ii) The *MAP threshold* as $\mathtt{h}^{\mathrm{MAP}} \triangleq$

$$\inf\left\{\mathtt{h} \in [0,1] \mid \liminf_{n\to\infty} \tfrac{1}{n}\mathbb{E}\left[\mathrm{H}\left(X^n \mid Y^n(\mathtt{c}(\mathtt{h}))\right)\right] > 0\right\},$$

where the expectation $\mathbb{E}[\cdot\,]$ is over the LDPC ensemble.

iii) The *potential threshold* as

$$\mathtt{h}^* \triangleq \sup\{\mathtt{h} \in [0,1] \mid \Delta E(\mathtt{c}(\mathtt{h})) > 0\}.$$

iv) The *stability threshold* as

$$\mathtt{h}^{\mathrm{stab}} \triangleq \sup\{\mathtt{h} \in [0,1] \mid \mathfrak{B}(\mathtt{c}(\mathtt{h}))\lambda'(0)\rho'(1) < 1\}.$$

In the sequel, the potential threshold and its role in connecting the BP and MAP thresholds are paramount. In particular, the region where $\Delta E(\mathtt{c}(\mathtt{h})) > 0$ characterizes the BP performance of the spatially-coupled ensemble, and, by definition of the potential threshold and Lemma 26(iii), if $\mathtt{h} < \mathtt{h}^*$, then $\Delta E(\mathtt{c}(\mathtt{h})) > 0$.

The stability threshold establishes an important technical property of the potential functional. When $\mathtt{h}^{\mathrm{stab}} = 1$, any constraints involving $\mathtt{h}^{\mathrm{stab}}$ are *superfluous*. For LDPC ensembles with no degree-two variable-nodes, $\mathtt{h}^{\mathrm{stab}} = 1$. For ensembles with degree-two variable-nodes[2], $0 < \mathtt{h}^{\mathrm{stab}} \leq 1$.

Lemma 29: The following properties regarding the stability threshold hold.

[2]We exclude ensembles with degree-one variable-nodes.

i) $\mathtt{h}^* \leq \mathtt{h}^{\text{stab}}$

ii) If $\mathtt{h} < \mathtt{h}^{\text{stab}}$, $\Delta_\infty \in (\mathcal{V}(\mathtt{c}(\mathtt{h})))^o$, the interior of the set $\mathcal{V}(\mathtt{c}(\mathtt{h}))$ in $(\mathcal{X}, d_{\text{H}})$.

Proof. See Section II.H.3.3. □

Lemma 30: If $\mathtt{h}^* < \mathtt{h}^{\text{stab}}$, then for $\mathtt{h} > \mathtt{h}^*$ there exists an $\mathtt{x} \in \mathcal{X}$ such that $U_{\text{s}}(\mathtt{x}; \mathtt{c}(\mathtt{h})) < 0$.

Proof. See Section II.H.3.4. □

Remark 31: Negativity of the potential functional beyond the potential threshold is important. This allows us to relate the potential and MAP threshold (Lemma 32). Negativity is also used in the converse of the threshold saturation result (Theorem 47). For a family of BEC or binary AWGN channels, Lemma 30 can be extended to include the case $\mathtt{h}^* = \mathtt{h}^{\text{stab}}$. We conjecture that this holds for *any* family of BMS channels. See Section II.H.6 for a further discussion.

Lemma 32: For an LDPC ensemble without odd-degree check-nodes over any BMS channel, or any LDPC ensemble over the BEC or the binary AWGN channel,

i) $\liminf_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\text{H}\left(X^n | Y^n(\mathtt{c}(\mathtt{h}))\right)\right] \geq -\inf_{\mathtt{x}\in\mathcal{X}} U_{\text{s}}(\mathtt{x}; \mathtt{c}(\mathtt{h}))$,

ii) If $\mathtt{h}^* < \mathtt{h}^{\text{stab}}$, then $\mathtt{h}^{\text{MAP}} \leq \mathtt{h}^*$.

Proof. i) Since the potential functional is the negative of the replica-symmetric free entropy calculated in [36, 39, 42], the main result of these papers translates directly into the desired result.

ii) Let $\mathtt{h} > \mathtt{h}^*$. Since $\mathtt{h}^* < \mathtt{h}^{\text{stab}}$ by assumption, from Lemma 30 and part i,

$$\liminf_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\text{H}\left(X^n | Y^n(\mathtt{c}(\mathtt{h}))\right)\right] \geq -\inf_{\mathtt{x}\in\mathcal{X}} U_{\text{s}}(\mathtt{x}; \mathtt{c}(\mathtt{h})) > 0.$$

Thus, by Definition 28(ii), $\mathtt{h} \geq \mathtt{h}^{\text{MAP}}$. Hence $\mathtt{h}^* \geq \mathtt{h}^{\text{MAP}}$.

□

The following remark discusses, rather informally, further connections between single and spatially-coupled system thresholds, based on results from [43], [19].

Remark 33: Let $\mathtt{h}_{\mathrm{c}}^{\mathrm{BP}}$ and $\mathtt{h}_{\mathrm{c}}^{\mathrm{MAP}}$ denote the BP and MAP thresholds, respectively, of the spatially-coupled system by first letting the chain length and then the coupling width go to infinity. Our results establish (Theorems 44 and 47) that

$$\mathtt{h}_{\mathrm{c}}^{\mathrm{BP}} = \mathtt{h}^{*}. \tag{II.2}$$

In [43] it is shown that, under some restrictions on the degree distributions[3], $\mathtt{h}_{\mathrm{c}}^{\mathrm{MAP}} = \mathtt{h}^{\mathrm{MAP}}$. By Lemma 32, for any ensemble with $\mathtt{h}^{\mathrm{stab}} = 1$, e.g. an ensemble with no degree-two variable nodes, $\mathtt{h}^{\mathrm{MAP}} \leq \mathtt{h}^{*}$. Combining these results with optimality of the MAP decoder and (II.2)

$$\mathtt{h}^{\mathrm{MAP}} \leq \mathtt{h}^{*} = \mathtt{h}_{\mathrm{c}}^{\mathrm{BP}} \leq \mathtt{h}_{\mathrm{c}}^{\mathrm{MAP}} = \mathtt{h}^{\mathrm{MAP}}.$$

This shows that $\mathtt{h}^{*} = \mathtt{h}^{\mathrm{MAP}}$, for an ensemble satisfying the aforementioned conditions.

The threshold saturation result shown in [19] can be summarized as follows. For regular codes with left-degree d_v, right-degree d_c, and a *smooth* family of channels, the BP threshold is equal to the *area* threshold $\mathtt{h}_{\mathrm{c}}^{\mathrm{BP}} = \mathtt{h}^{A}$, where the area threshold is

$$\mathtt{h}^{A} \triangleq \sup\left\{\mathtt{h} \in [0,1] \mid A(\mathsf{T}_{\mathrm{s}}^{(\infty)}(\Delta_0; \mathtt{c}(\mathtt{h})), d_v, d_c) \leq 0\right\},$$

and

$$A(\mathtt{x}, d_v, d_c) \triangleq \mathrm{H}\left(\mathtt{x}\right) + \Big(d_v - 1 - \frac{d_v}{d_c}\Big)\mathrm{H}\left(\mathtt{x}^{\boxplus d_c}\right) - (d_v - 1)\mathrm{H}\left(\mathtt{x}^{\boxplus d_c - 1}\right).$$

At the DE fixed point $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\Delta_0; \mathtt{c}(\mathtt{h}))$, using the duality rule for entropy (Proposition 4), it is also easy to show that

$$A(\mathsf{T}_{\mathrm{s}}^{(\infty)}(\Delta_0; \mathtt{c}(\mathtt{h})), d_v, d_c) = -U_{\mathrm{s}}(\mathsf{T}_{\mathrm{s}}^{(\infty)}(\Delta_0; \mathtt{c}(\mathtt{h})); \mathtt{c}(\mathtt{h})).$$

This immediately implies that $\mathtt{h}^{*} \leq \mathtt{h}^{A}$. Therefore, by [19, Theorem 41], $\mathtt{h}^{A} = \mathtt{h}_{\mathrm{c}}^{\mathrm{BP}}$, and the results here, (II.2), $\mathtt{h}^{*} = \mathtt{h}^{A}$. Hence, the thresholds $\mathtt{h}^{\mathrm{MAP}}$, $\mathtt{h}^{*}$ and $\mathtt{h}^{A}$ are all equal under suitable conditions.

In particular, for regular codes with even-degree checks, it has been shown rigor-

[3]Requires regular check-nodes with even degree; this can be relaxed to $R(t)$ convex on $[-1, 1]$.

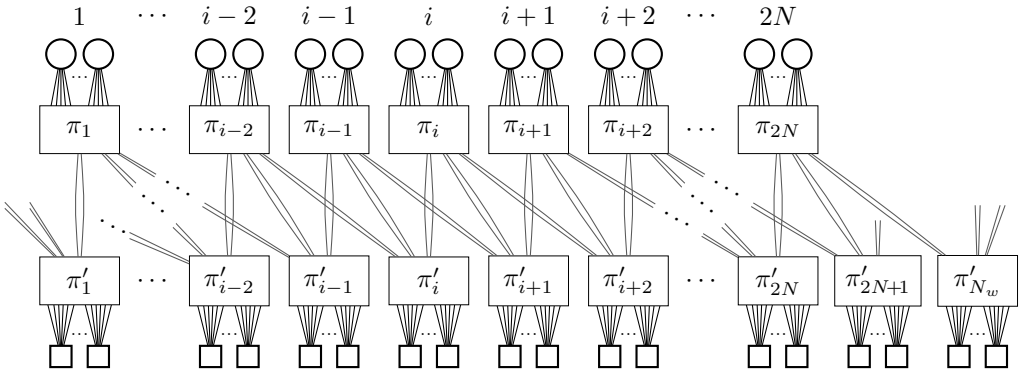

Figure II.2: An example of a $(\lambda(t) = t^4, \rho(t) = t^5, N, w = 3)$ spatially-coupled LDPC ensemble. Sockets in each variable- and check-node group are permuted (π and π' denote the permutations) and partitioned into w groups, and connected as shown above. This results in some sockets of the check-node groups at the boundary unconnected.

ously that $\mathtt{h}^{\mathrm{MAP}} = \mathtt{h}^{A}$. However, it is instructive to note that the Maxwell conjecture [44, Conjecture 1], which states that the MAP GEXIT function is obtained by applying the Maxwell construction to the EBP GEXIT curve, is yet to be established for BMS channels.

II.C.2 Coupled System

The potential theory for single systems is now extended to spatially-coupled systems. Vectors of measures are denoted by underlines (e.g., $\underline{\mathsf{x}}$) with $[\underline{\mathsf{x}}]_i = \mathsf{x}_i$. Functionals operating on a single measure are distinguished from those operating on vectors by their input (i.e., $F(\mathsf{x})$ vs. $F(\underline{\mathsf{x}})$). Also, for vectors $\underline{\mathsf{x}}'$ and $\underline{\mathsf{x}}$, we write $\underline{\mathsf{x}}' \succeq \underline{\mathsf{x}}$ if $\mathsf{x}'_i \succeq \mathsf{x}_i$ for all i, and $\underline{\mathsf{x}}' \succ \underline{\mathsf{x}}$ if $\underline{\mathsf{x}}'_i \succeq \underline{\mathsf{x}}_i$ for all i and $\mathsf{x}'_i \succ \mathsf{x}_i$ for some i.

The ideas underlying spatial coupling now appear to be quite general. The local coupling in the system allows the effect of the perfect information, provided at the boundary, to propagate throughout the system. In the large-system limit, these coupled systems show a significant performance improvement. The spatially-coupled system model is now described.

The (λ, ρ, N, w) spatially-coupled LDPC ensemble is defined as follows. As before, the node perspective degree distributions are denoted by L, R, and

$$L(t) = \sum_{n=0}^{\deg(L)} L_n t^n, \qquad\qquad R(t) = \sum_{n=0}^{\deg(R)} R_n t^n.$$

A collection of $2N$ variable-node groups are placed at all positions in $\mathcal{N}_v = \{1, 2, \ldots, 2N\}$ and a collection of $2N + (w-1)$ check-node groups are placed at all positions in $\mathcal{N}_c = \{1, 2, \ldots, 2N + (w-1)\}$. For notational convenience, the rightmost check-node group index is denoted by $N_w \triangleq 2N + (w-1)$. For the below construction of a spatially-coupled LDPC ensemble, we assume all L_n, R_n are rational.

The integer M is chosen large enough so that i) ML_i, $ML'(1)R_j/R'(1)$ are natural numbers for $1 \leq i \leq \deg(L)$, $1 \leq j \leq \deg(R)$, and ii) $ML'(1)$ is divisible by w. At each variable-node group, ML_i nodes of degree i are placed for $1 \leq i \leq \deg(L)$. Similarly, at each check-node group, $ML'(1)R_j/R'(1)$ nodes of degree j are placed for $1 \leq j \leq \deg(R)$. At each variable-node and check-node group, the $ML'(1)$ edge sockets are partitioned into w equal-sized groups using a uniform random permutation. Denote these partitions, respectively, by $\mathcal{P}^v_{i,k}$ and $\mathcal{P}^c_{j,k}$ at variable-node and check-node groups, where $1 \leq i \leq 2N$, $1 \leq j \leq N_w$ and $1 \leq k \leq w$. The spatially-coupled system is constructed by connecting the sockets in $\mathcal{P}^v_{i,k}$ to sockets in $\mathcal{P}^c_{i+k-1,k}$ using uniform random permutations. This construction leaves some sockets of the check-node groups at the boundaries unconnected and these sockets are assigned the binary value 0 (i.e., the socket and edge are removed). These 0 values form the perfect information that gets decoding started. A Tanner graph example of a spatially-coupled LDPC ensemble depicting these connections is provided in Fig. II.2.

The analysis below is valid for any spatially-coupled system whose density evolution is given by (II.4). For the random ensemble described in [10, Section II-B], and for the (λ, ρ, N, w) ensemble described above, the asymptotic density evolution is indeed described by (II.4). Thus, our analysis holds for both these ensembles. However, this is no longer true for the protograph construction described in [10, Section II-A].

Let $\tilde{\mathsf{x}}_i^{(\ell)}$ be the variable-node output distribution at node i after ℓ iterations of message passing. Then, the input distribution to the i-th check-node group is the normalized sum of averaged variable-node output distributions,

$$\mathsf{x}_i^{(\ell)} = \frac{1}{w} \sum_{k=0}^{w-1} \tilde{\mathsf{x}}_{i-k}^{(\ell)}. \tag{II.3}$$

The averaging in the reversed direction (i.e. from check-node to the variable-node) follows naturally from this setup and is essentially the transpose of the forward averaging for the check-node output distributions. This model uses uniform coupling

over a fixed window, but in a more general setting window size and coefficient weights could vary from node to node. By virtue of the fixed boundary condition, $\tilde{\mathsf{x}}_i^{(\ell)} = \Delta_\infty$ for $i \notin \mathcal{N}_v$ and all ℓ, and from the relation in (II.3), this implies $\mathsf{x}_i^{(\ell)} = \Delta_\infty$ for $i \notin \mathcal{N}_c$ and all ℓ.

Generalizing [19, Eqn. 12] to irregular codes gives the evolution of the variable-node output distributions,

$$\tilde{\mathsf{x}}_i^{(\ell+1)} = \mathsf{c} \circledast \lambda^{\circledast}\left(\frac{1}{w}\sum_{j=0}^{w-1} \rho^{\boxplus}\left(\frac{1}{w}\sum_{k=0}^{w-1} \tilde{\mathsf{x}}_{i+j-k}^{(\ell)}\right)\right). \tag{II.4}$$

Making a change of variables, the variable-node output distribution evolution in (II.4) can be rewritten in terms of check-node input distributions

$$\mathsf{x}_i^{(\ell+1)} = \frac{1}{w}\sum_{k=0}^{w-1} \mathsf{c}_{i-k} \circledast \lambda^{\circledast}\left(\frac{1}{w}\sum_{j=0}^{w-1} \rho^{\boxplus}\left(\mathsf{x}_{i-k+j}^{(\ell)}\right)\right), \tag{II.5}$$

for $i \in \mathcal{N}_c$, where $\mathsf{c}_i = \mathsf{c}$ when $i \in \mathcal{N}_v$ and $\mathsf{c}_i = \Delta_\infty$ otherwise. While (II.4) is a more natural representation for the underlying system, (II.5) is more mathematically tractable and easily yields a coupled potential functional. As such, we adopt the system characterized by (II.5) and refer to it as the (λ, ρ, N, w) *spatially-coupled LDPC system*.

Borrowing notation from the single system, when the spatially-coupled system with channel c is initialized with $\underline{\mathsf{a}}$ (i.e. $\mathsf{x}_i^{(0)} = \mathsf{a}_i$), the check-node input distribution after ℓ iterations of message-passing is denoted by $\mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}}; \mathsf{c})$. One iteration of this message-passing is also denoted by $\mathsf{T}_{\mathrm{c}}(\underline{\mathsf{a}}; \mathsf{c})$. With this new notation, (II.5) can be written compactly as

$$\mathsf{x}_i^{(\ell+1)} = \mathsf{T}_{\mathrm{c}}(\underline{\mathsf{x}}^{(\ell)}; \mathsf{c})_i.$$

If the sequence of measure vectors $\{\mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}}; \mathsf{c})\}_{\ell=1}^{\infty}$ converges pointwise, then its limit is denoted by $\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}}; \mathsf{c})$. The following proposition establishes certain monotonicity properties of $\mathsf{T}_{\mathrm{c}}^{(\ell)}$.

Lemma 34: The operator $\mathsf{T}_{\mathrm{c}}^{(\ell)} \colon \mathcal{X}^{N_w} \times \mathcal{X} \to \mathcal{X}^{N_w}$ satisfies the following for all $1 \leq \ell < \infty$.

i) If $\underline{\mathsf{a}}_1 \succeq \underline{\mathsf{a}}_2$, then $\mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}}_1; \mathsf{c}) \succeq \mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}}_2; \mathsf{c})$ for all $\mathsf{c} \in \mathcal{X}$.

ii) If $\mathsf{c}_1 \succeq \mathsf{c}_2$, then $\mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}};\mathsf{c}_1) \succeq \mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}};\mathsf{c}_2)$ for all $\underline{\mathsf{a}} \in \mathcal{X}^{N_w}$.

iii) If $\mathsf{T}_{\mathrm{c}}(\underline{\mathsf{a}};\mathsf{c}) \preceq \underline{\mathsf{a}}$, then $\mathsf{T}_{\mathrm{c}}^{(\ell+1)}(\underline{\mathsf{a}};\mathsf{c}) \preceq \mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}};\mathsf{c})$. Also, the limit $\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}};\mathsf{c})$ exists and satisfies $\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}};\mathsf{c}) \preceq \mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}};\mathsf{c})$,

$$\mathsf{T}_{\mathrm{c}}(\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}};\mathsf{c});\mathsf{c}) = \mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}};\mathsf{c}).$$

iv) If $\mathsf{T}_{\mathrm{c}}(\underline{\mathsf{a}};\mathsf{c}) \succeq \underline{\mathsf{a}}$, then $\mathsf{T}_{\mathrm{c}}^{(\ell+1)}(\underline{\mathsf{a}};\mathsf{c}) \succeq \mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}};\mathsf{c})$. Also, the limit $\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}};\mathsf{c})$ exists and satisfies $\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}};\mathsf{c}) \succeq \mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}};\mathsf{c})$,

$$\mathsf{T}_{\mathrm{c}}(\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}};\mathsf{c});\mathsf{c}) = \mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}};\mathsf{c}).$$

Proof. The proof is almost identical to the proof of Lemma 18. We skip the details for brevity. □

When the spatially-coupled system is initialized with

$$\mathsf{x}_i^{(0)} = \Delta_0, \quad 1 \le i \le N_w,$$

the uniform coupling coefficients and symmetric boundary conditions induce left-right symmetry on $\underline{\mathsf{x}}^{(\ell)}$. In particular, the spatially-coupled system is fully described by only half the distributions because

$$\mathsf{x}_i^{(\ell)} = \mathsf{x}_{2N+w-i}^{(\ell)},$$

for all ℓ. As density evolution progresses, the perfect information from the boundary propagates inward. This propagation induces a non-decreasing degradation ordering on positions $1, \ldots, \lceil N_w/2 \rceil$ and a non-increasing degradation ordering on positions $\lceil N_w/2 \rceil + 1, \ldots, N_w$. For example, see Fig. II.3.

This ordering introduces a degraded maximum at $i_0 \triangleq N + \lceil \frac{w-1}{2} \rceil$, and this maximum allows one to define a modified recursion that upper bounds the spatially-coupled system.

Definition 35: The *modified system* is a modification of (II.5) defined by fixing the values of positions outside $\mathcal{N}_c' \triangleq \{1, 2, \ldots, i_0\}$, where i_0 is defined as above. As before, the boundary is fixed to Δ_∞, that is $\mathsf{x}_i^{(\ell)} = \Delta_\infty$ for $i \notin \mathcal{N}_c$ and all ℓ. More importantly, it fixes the values $\mathsf{x}_i^{(\ell)} = \mathsf{x}_{i_0}^{(\ell)}$ for $i_0 < i \le N_w$ and all ℓ.

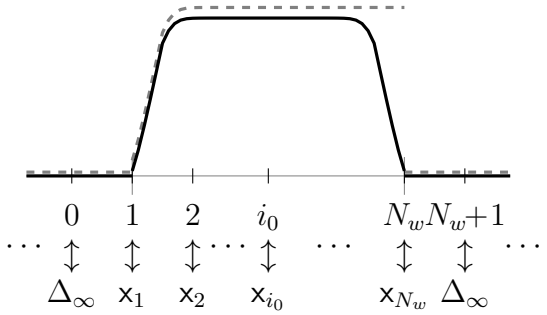

Figure II.3: This figure depicts the entropies of $\mathsf{x}_1, \cdots, \mathsf{x}_{N_w}$ in a typical iteration. The solid line corresponds to the spatially-coupled system and the dashed line to the modified system. The distributions of the modified system are always degraded with respect to the spatially-coupled system, hence a higher entropy. The distributions outside the set $\{1, \cdots, N_w\}$ are fixed to Δ_∞ for both the systems.

The DE update of the modified system is identical to (II.5) for the first i_0 terms, $1, \ldots, i_0$, but a secondary update is required to impose the saturation constraint, $\mathsf{x}_i = \mathsf{x}_{i_0}$ for $i_0 < i \leq N_w$. Repeated iterations for this system require that this saturation constraint is applied at every step. The distributions of modified system are degraded with respect to that of spatially-coupled system, thus the modified system serves as a convenient upper bound for the spatially-coupled system. Both the spatially-coupled system and the modified system are collectively referred to as *coupled systems*.

In Fig. II.3, the entropies of the two systems are illustrated in a typical iteration. We emphasize that the operator T_{c} refers to the spatially-coupled system, *not* the modified system. However, the DE update for the modified system also satisfies the same monotonicity properties of T_{c} in Lemma 34.

If either spatially-coupled system or modified system is initialized with $\underline{\mathsf{x}}^{(0)} = \underline{\Delta_0} \triangleq \{\Delta_0, \ldots, \Delta_0\}$, then the sequence of measure vectors $\{\underline{\mathsf{x}}^{(\ell)}\}$, by Lemma 34, satisfies $\underline{\mathsf{x}}^{(\ell+1)} \preceq \underline{\mathsf{x}}^{(\ell)}$ and converges to a fixed point $\underline{\mathsf{x}}$. Thus, for the spatially-coupled system,

$$\underline{\mathsf{x}} = \mathsf{T}_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c}).$$

Such a fixed point for the modified system satisfies an additional property, stated in the following lemma.

Lemma 36: The fixed point $\underline{\mathsf{x}}$ resulting from initializing the modified system with

$\underline{\Delta_0}$ satisfies

$$\mathsf{x}_i \succeq \mathsf{x}_{i-1}, \quad 2 \leq i \leq N_w$$

Proof. See Section II.H.3.5. □

Now, we define the coupled potential. The definitions below pertain to both spatially-coupled and modified system.

Definition 37: The coupled potential functional $U_{\mathrm{c}} \colon \mathcal{X}^{N_w} \times \mathcal{X} \to \mathbb{R}$ is given by

$$\begin{aligned} U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c}) \triangleq L'(1) \sum_{i=1}^{N_w} \left[\frac{1}{R'(1)} \mathrm{H}\left(R^{\circledast}(\mathsf{x}_i)\right) + \mathrm{H}\left(\rho^{\circledast}(\mathsf{x}_i)\right) - \mathrm{H}\left(\mathsf{x}_i \circledast \rho^{\circledast}(\mathsf{x}_i)\right) \right] \\ - \sum_{i=1}^{2N} \mathrm{H}\left(\mathsf{c} \circledast L^{\circledast}\Big(\frac{1}{w} \sum_{j=0}^{w-1} \rho^{\circledast}(\mathsf{x}_{i+j}) \Big) \right). \end{aligned} \tag{II.6}$$

Lemma 38: The directional derivative of the potential functional in (II.6) with respect to $\underline{\mathsf{x}} \in \mathcal{X}^{N_w}$, evaluated in the direction $\underline{\mathsf{y}} \in \mathcal{X}_{\mathrm{d}}^{N_w}$ is given by

$$\mathrm{d}_{\underline{\mathsf{x}}} U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c})[\underline{\mathsf{y}}] = L'(1) \sum_{i=1}^{N_w} \mathrm{H}\left((\mathsf{T}_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c})_i - \mathsf{x}_i) \circledast \rho'^{\circledast}(\mathsf{x}_i) \circledast \mathsf{y}_i \right). \tag{II.7}$$

Proof. See Section II.H.3.6. □

Lemma 39: The second-order directional derivative of the potential functional in (II.6) with respect to $\underline{\mathsf{x}}$, evaluated in the direction $[\underline{\mathsf{y}}, \underline{\mathsf{z}}] \in \mathcal{X}_{\mathrm{d}}^{N_w} \times \mathcal{X}_{\mathrm{d}}^{N_w}$ is given by

$$\begin{aligned} &\mathrm{d}^2_{\underline{\mathsf{x}}} U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c})[\underline{\mathsf{y}}, \underline{\mathsf{z}}] = \\ &L'(1) \sum_{i=1}^{N_w} \left[\rho''(1) \mathrm{H}\left(\mathsf{T}_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c})_i \circledast \frac{\rho''^{\circledast}(\mathsf{x}_i)}{\rho''(1)} \circledast \mathsf{y}_i \circledast \mathsf{z}_i \right) - \rho''(1) \mathrm{H}\left(\mathsf{x}_i \circledast \frac{\rho''^{\circledast}(\mathsf{x}_i)}{\rho''(1)} \circledast \mathsf{y}_i \circledast \mathsf{z}_i \right) \right] \\ &- L'(1) \sum_{i=1}^{N_w} \rho'(1) \mathrm{H}\left(\frac{\rho'^{\circledast}(\mathsf{x}_i)}{\rho'(1)} \circledast \mathsf{y}_i \circledast \mathsf{z}_i \right) - \frac{L'(1)\lambda'(1)\rho'(1)^2}{w} \sum_{i=1}^{N_w} \sum_{m=\max\{i-(w-1),1\}}^{\min\{i+(w-1),N_w\}} \cdots \\ &\mathrm{H}\left(\frac{1}{w} \sum_{k=0}^{w-1} \mathsf{c}_{i-k} \circledast \frac{\lambda'^{\circledast}\left(\frac{1}{w} \sum_{j=0}^{w-1} \rho^{\circledast}(\mathsf{x}_{i-k+j}) \right)}{\lambda'(1)} \circledast \left[\frac{\rho'^{\circledast}(\mathsf{x}_i)}{\rho'(1)} \circledast \mathsf{z}_i \right] \circledast \left[\frac{\rho'^{\circledast}(\mathsf{x}_m)}{\rho'(1)} \circledast \mathsf{y}_m \right] \right) \end{aligned} \tag{II.8}$$

Proof. See Section II.H.3.7. □

II.D THRESHOLD SATURATION FOR LDPC ENSEMBLES

II.D.1 Achievability of Threshold Saturation

We now prove threshold saturation for spatially-coupled LDPC ensembles. For a family of BMS channels, we will show that, if $\mathtt{h} < \mathtt{h}^*$, then the only fixed point of the modified system is $\underline{\Delta_\infty}$. Since the modified system is an upper bound on the spatially-coupled system, we then conclude that the only fixed point of the spatially-coupled system is $\underline{\Delta_\infty}$.

Consider a modified system with potential functional U_{c} as in Definition 37, and a non-trivial fixed point $\underline{\mathsf{x}}$. Also, consider a parameterization $\phi\colon [0,1] \to \mathbb{R}$, where

$$\phi(t) = U_{\mathrm{c}}(\underline{\mathsf{x}} + t(\underline{\mathsf{x}}' - \underline{\mathsf{x}}); \mathsf{c}(\mathtt{h})).$$

The path endpoint $\underline{\mathsf{x}}'$ is chosen to be a small perturbation of $\underline{\mathsf{x}}$. For all channels $\mathsf{c}(\mathtt{h})$ with $\mathtt{h} < \mathtt{h}^*$, at $\underline{\mathsf{x}}$, it can be shown that the potential functional decreases, at least by a constant independent of the modified system, along the perturbation $\underline{\mathsf{x}}'$. Moreover, a fixed point is also a stationary point of the potential functional. Also, at the fixed point, the second-order variations in the potential can be made arbitrarily small by choosing a large coupling parameter w. Thus, all variations in the potential functional up to second-order can be made arbitrarily small.

By calculating the change in potential at a non-trivial fixed point in two different ways: first by explicit calculation of change in the potential and second by the first- and second-order variations, one obtains a contradiction to the existence of a non-trivial fixed point from the second-order Taylor expansion of $\phi(t)$, for all $\mathsf{c}(\mathtt{h})$ with $\mathtt{h} < \mathtt{h}^*$.

These ideas are formalized below. A right shift is chosen for the perturbation and the shift operator $\mathsf{S}(\cdot)$ is defined in Definition 40. In Lemma 41, we bound the change in potential due to shift. Lemmas 42 and 43 characterize the first- and second-order variations, respectively, along the shift direction $[\mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}]$, for a non-trivial fixed point $\underline{\mathsf{x}}$. Finally, Theorem 44 proves threshold saturation.

Definition 40: The shift operator $\mathsf{S}\colon \mathcal{X}^{N_w} \to \mathcal{X}^{N_w}$ is defined pointwise by

$$[\mathsf{S}(\underline{\mathsf{x}})]_1 \triangleq \Delta_\infty, \qquad\qquad [\mathsf{S}(\underline{\mathsf{x}})]_i \triangleq \mathsf{x}_{i-1}, \quad 2 \le i \le N_w.$$

Lemma 41: Let $\underline{\mathsf{x}} \in \mathcal{X}^{N_w}$ be such that $\mathsf{x}_i = \mathsf{x}_{i_0}$, for $i_0 \leq i \leq N_w$. Then the change in the potential functional for a modified system associated with the shift operator is bounded by

$$U_{\mathrm{c}}(\mathsf{S}(\underline{\mathsf{x}}); \mathsf{c}) - U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c}) \leq -U_{\mathrm{s}}(\mathsf{x}_{i_0}; \mathsf{c}).$$

Proof. See Section II.H.4.1. □

Lemma 42: If $\underline{\mathsf{x}} \succ \underline{\Delta_\infty} \triangleq [\Delta_\infty, \ldots, \Delta_\infty]$ is a fixed point of the modified system resulting from $\underline{\Delta_0}$ initialization, then

$$\mathrm{d}_{\underline{\mathsf{x}}} U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c})[\mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}] = 0,$$

and moreover x_{i_0} is not in the basin of attraction to Δ_∞ (i.e., $\mathsf{x}_{i_0} \notin \mathcal{V}(\mathsf{c})$).

Proof. See Section II.H.4.2. □

The above two lemmas together with Definition 25(ii) imply that for a non-trivial fixed point $\underline{\mathsf{x}}$ resulting from initializing the modified system with $\underline{\Delta_0}$,

$$U_{\mathrm{c}}(\mathsf{S}(\underline{\mathsf{x}}); \mathsf{c}) - U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c}) \leq -U_{\mathrm{s}}(\mathsf{x}_{i_0}; \mathsf{c}) \leq -\Delta E(\mathsf{c}).$$

Thus, when $\Delta E(\mathsf{c}) > 0$, the absolute change in potential due to shift is lower bounded by a constant independent of $\underline{\mathsf{x}}$, N, w, and hence of the coupled system.

Lemma 43: Suppose $\underline{\mathsf{x}}$ is a fixed point of the modified system resulting from $\underline{\Delta_0}$ initialization. The second-order directional derivative of $U_{\mathrm{c}}(\underline{\mathsf{x}}_1; \mathsf{c})$ with respect to $\underline{\mathsf{x}}_1$, evaluated along $[\mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}, \mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}]$, can be absolutely bounded with

$$\left| \mathrm{d}^2_{\underline{\mathsf{x}}_1} U_{\mathrm{c}}(\underline{\mathsf{x}}_1; \mathsf{c})[\mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}, \mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}] \right| \leq \frac{K_{\lambda,\rho}}{w},$$

where the constant $K_{\lambda,\rho} \triangleq L'(1)\left(2\rho''(1) + \rho'(1) + 2\lambda'(1)\rho'(1)^2\right)$ is independent of N and w.

Proof. See Section II.H.4.3. □

Theorem 44: Fix a family of BMS channels $\mathsf{c}(\mathtt{h})$, and the LDPC(λ, ρ) ensemble. For $\mathtt{h} < \mathtt{h}^*$, all N, and any $w > K_{\lambda,\rho}/(2\Delta E(\mathsf{c}(\mathtt{h})))$, the only fixed point of density

evolution for the spatially-coupled LDPC (λ, ρ, N, w) ensemble with channel $\mathtt{c}(\mathtt{h})$ is $\underline{\Delta_\infty}$.

Proof. First, since $\mathtt{h} < \mathtt{h}^*$, $\Delta E(\mathtt{c}(\mathtt{h})) > 0$. Consider a modified system with a fixed $w > K_{\lambda,\rho}/(2\Delta E(\mathtt{c}(\mathtt{h})))$ and any N. Suppose $\underline{\mathtt{x}}$ is a fixed point of modified system resulting from $\underline{\Delta_0}$ initialization. If $\underline{\mathtt{x}} = \underline{\Delta_\infty}$, by the monotonicity of the DE update resulting from $\underline{\Delta_0}$ initialization, there is no other fixed point for the modified system. Suppose instead that $\underline{\mathtt{x}} \succ \underline{\Delta_\infty}$. In this case, we will arrive at a contradiction in the following.

Let $\underline{\mathtt{y}} = \mathsf{S}(\underline{\mathtt{x}}) - \underline{\mathtt{x}}$ and define $\phi\colon [0,1] \to \mathbb{R}$ by

$$\phi(t) = U_{\mathrm{c}}(\underline{\mathtt{x}} + t\underline{\mathtt{y}}; \mathtt{c}(\mathtt{h})).$$

This is well defined because, for all $t \in [0,1]$, $\underline{\mathtt{x}} + t\underline{\mathtt{y}} = (1-t)\underline{\mathtt{x}} + t\mathsf{S}(\underline{\mathtt{x}})$ is a vector of probability measures. As in Proposition 16, ϕ is a polynomial in t, and thus infinitely differentiable over the entire unit interval. Hence, the second-order Taylor series expansion about $t = 0$, evaluated at $t = 1$, provides

$$\phi(1) = \phi(0) + \phi'(0)(1-0) + \tfrac{1}{2}\phi''(t_0)(1-0)^2, \tag{II.9}$$

for some $t_0 \in [0,1]$. The first and second derivatives of ϕ are characterized by the first- and second-order directional derivatives of U_{c}:

$$\begin{aligned}\phi'(t) &= \lim_{\delta \to 0} \frac{U_{\mathrm{c}}(\underline{\mathtt{x}} + (t+\delta)\underline{\mathtt{y}}; \mathtt{c}(\mathtt{h})) - U_{\mathrm{c}}(\underline{\mathtt{x}} + t\underline{\mathtt{y}}; \mathtt{c}(\mathtt{h}))}{\delta} \\ &= \mathrm{d}_{\underline{\mathtt{x}}_1} U_{\mathrm{c}}(\underline{\mathtt{x}}_1; \mathtt{c}(\mathtt{h}))[\underline{\mathtt{y}}] \Big|_{\underline{\mathtt{x}}_1 = \underline{\mathtt{x}} + t\underline{\mathtt{y}}},\end{aligned}$$

and similarly,

$$\phi''(t) = \mathrm{d}^2_{\underline{\mathtt{x}}_1} U_{\mathrm{c}}(\underline{\mathtt{x}}_1; \mathtt{c}(\mathtt{h}))[\underline{\mathtt{y}}, \underline{\mathtt{y}}] \Big|_{\underline{\mathtt{x}}_1 = \underline{\mathtt{x}} + t\underline{\mathtt{y}}}.$$

Substituting and rearranging terms in (II.9) provides

$$\begin{aligned}\tfrac{1}{2}\mathrm{d}^2_{\underline{\mathtt{x}}_1} U_{\mathrm{c}}(\underline{\mathtt{x}}_1; \mathtt{c}(\mathtt{h}))[\underline{\mathtt{y}}, \underline{\mathtt{y}}] \Big|_{\underline{\mathtt{x}}_1 = \underline{\mathtt{x}} + t_0\underline{\mathtt{y}}} &= U_{\mathrm{c}}(\mathsf{S}(\underline{\mathtt{x}}); \mathtt{c}(\mathtt{h})) - U_{\mathrm{c}}(\underline{\mathtt{x}}; \mathtt{c}(\mathtt{h})) - \mathrm{d}_{\underline{\mathtt{x}}} U_{\mathrm{c}}(\underline{\mathtt{x}}; \mathtt{c}(\mathtt{h}))[\mathsf{S}(\underline{\mathtt{x}}) - \underline{\mathtt{x}}] \\ &= U_{\mathrm{c}}(\mathsf{S}(\underline{\mathtt{x}}); \mathtt{c}(\mathtt{h})) - U_{\mathrm{c}}(\underline{\mathtt{x}}; \mathtt{c}(\mathtt{h})) \quad \text{(Lemma 42)} \\ &\leq -U_{\mathrm{s}}(\mathtt{x}_{i_0}; \mathtt{c}) \quad \text{(Lemma 41)}\end{aligned}$$

$$\leq -\Delta E(\mathtt{c}(\mathtt{h})). \quad \text{(Lemma 42 and Definition 25(ii))}$$

Taking the absolute value and applying the second order directional derivative bound from Lemma 43 gives

$$\Delta E(\mathtt{c}(\mathtt{h})) \leq \frac{K_{\lambda,\rho}}{2w} \quad \Longrightarrow \quad w \leq \frac{K_{\lambda,\rho}}{2\Delta E(\mathtt{c}(\mathtt{h}))},$$

a contradiction. Hence the only fixed point of the modified system is $\underline{\Delta_\infty}$. The distributions of the modified system are degraded with respect to the spatially-coupled system, and therefore, the only fixed point of the spatially-coupled system is also $\underline{\Delta_\infty}$. □

As an immediate consequence, for the (λ, ρ, N, w) spatially-coupled ensemble with $0 < K_{\lambda,\rho}/(2\Delta E(\mathtt{c}(\mathtt{h}))) < w < \infty$ and any N, its BP threshold is at least $\mathtt{h}$. Therefore, the BP threshold of the (λ, ρ, N, w) spatially-coupled ensemble, by first taking the limit $N \to \infty$ and then $w \to \infty$, is at least $\mathtt{h}^*$. Below, Theorem 47 establishes that, under $\mathtt{h}^* < \mathtt{h}^{\text{stab}}$, the BP threshold of the spatially-coupled ensemble in the limits given above is at most $\mathtt{h}^*$, which establishes the equality of the BP threshold to $\mathtt{h}^*$ in the above limits.

II.D.2 Converse to Threshold Saturation

We begin by establishing two monotonicity results.

Lemma 45: Consider $\underline{\mathtt{x}}_1 \in \mathcal{X}^{N_w}$ and $\underline{\mathtt{x}}_2 = \mathsf{T}_{\mathrm{c}}(\underline{\mathtt{x}}_1; \mathtt{c})$.

i) If $\underline{\mathtt{x}}_2 \succeq \underline{\mathtt{x}}_1$, $\mathsf{T}_{\mathrm{c}}(\underline{\mathtt{x}}_1 + t(\underline{\mathtt{x}}_2 - \underline{\mathtt{x}}_1); \mathtt{c}) \succeq \underline{\mathtt{x}}_1 + t(\underline{\mathtt{x}}_2 - \underline{\mathtt{x}}_1)$.

ii) If $\underline{\mathtt{x}}_2 \preceq \underline{\mathtt{x}}_1$, $\mathsf{T}_{\mathrm{c}}(\underline{\mathtt{x}}_1 + t(\underline{\mathtt{x}}_2 - \underline{\mathtt{x}}_1); \mathtt{c}) \preceq \underline{\mathtt{x}}_1 + t(\underline{\mathtt{x}}_2 - \underline{\mathtt{x}}_1)$.

Proof. i) If $\underline{\mathtt{x}}_2 \succeq \underline{\mathtt{x}}_1$, then for all $0 \leq t \leq 1$, $\underline{\mathtt{x}}_2 \succeq \underline{\mathtt{x}}_1 + t(\underline{\mathtt{x}}_2 - \underline{\mathtt{x}}_1) \succeq \underline{\mathtt{x}}_1$. Since T_{c} is order-preserving by Lemma 34,

$$\mathsf{T}_{\mathrm{c}}(\underline{\mathtt{x}}_1 + t(\underline{\mathtt{x}}_2 - \underline{\mathtt{x}}_1); \mathtt{c}) \succeq \mathsf{T}_{\mathrm{c}}(\underline{\mathtt{x}}_1; \mathtt{c}) = \underline{\mathtt{x}}_2 \succeq \underline{\mathtt{x}}_1 + t(\underline{\mathtt{x}}_2 - \underline{\mathtt{x}}_1).$$

ii) Follows by symmetry.

□

Lemma 46: Let $\underline{\mathtt{x}}_1 \in \mathcal{X}^{N_w}$, $\underline{\mathtt{x}}_2 = \mathsf{T}_{\mathrm{c}}(\underline{\mathtt{x}}_1; \mathtt{c})$, and suppose $\underline{\mathtt{x}}_2 \succeq \underline{\mathtt{x}}_1$ or $\underline{\mathtt{x}}_2 \preceq \underline{\mathtt{x}}_1$, then $U_{\mathrm{c}}(\underline{\mathtt{x}}_2; \mathtt{c}) \leq U_{\mathrm{c}}(\underline{\mathtt{x}}_1; \mathtt{c})$.

Proof. Assume $\underline{\mathsf{x}}_2 \succeq \underline{\mathsf{x}}_1$. Let $\phi\colon [0,1] \to \mathbb{R}$ be defined by

$$\phi(t) = U_{\mathrm{c}}(\underline{\mathsf{x}}_1 + t(\underline{\mathsf{x}}_2 - \underline{\mathsf{x}}_1); \mathsf{c}).$$

Observe that ϕ is a polynomial in t as in Proposition 16, with $\phi(0) = U_{\mathrm{c}}(\underline{\mathsf{x}}_1; \mathsf{c})$ and $\phi(1) = U_{\mathrm{c}}(\underline{\mathsf{x}}_2; \mathsf{c})$. Moreover,

$$\phi'(t) = \mathrm{d}_{\underline{\mathsf{x}}} U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c})[\underline{\mathsf{x}}_2 - \underline{\mathsf{x}}_1] \Big|_{\underline{\mathsf{x}} = \underline{\mathsf{x}}_1 + t(\underline{\mathsf{x}}_2 - \underline{\mathsf{x}}_1)}. \tag{II.10}$$

By Lemma 45,

$$\mathsf{T}_{\mathrm{c}}\left(\underline{\mathsf{x}}_1 + t(\underline{\mathsf{x}}_2 - \underline{\mathsf{x}}_1); \mathsf{c}\right) \succeq \underline{\mathsf{x}}_1 + t(\underline{\mathsf{x}}_2 - \underline{\mathsf{x}}_1),$$

and observing (II.7), the derivative in (II.10) is a sum of terms of the form

$$L'(1)\mathrm{H}\left([\mathsf{x}_3' - \mathsf{x}_3] \circledast \mathsf{x}_4 \circledast [\mathsf{x}_5' - \mathsf{x}_5]\right),$$

where $\mathsf{x}_3' \succeq \mathsf{x}_3$ and $\mathsf{x}_5' \succeq \mathsf{x}_5$, which is negative by Proposition 8(iii). For the case $\underline{\mathsf{x}}_2 \preceq \underline{\mathsf{x}}_1$, we can write a similar expression with $\mathsf{x}_3' \preceq \mathsf{x}_3$ and $\mathsf{x}_5' \preceq \mathsf{x}_5$. In either case, $\phi'(t) \leq 0$ for all $t \in [0,1]$. Thus, $U_{\mathrm{c}}(\underline{\mathsf{x}}_2; \mathsf{c}) = \phi(1) \leq \phi(0) = U_{\mathrm{c}}(\underline{\mathsf{x}}_1; \mathsf{c})$. □

Theorem 47: Fix a family of BMS channels $\mathsf{c}(\mathtt{h})$ and the LDPC(λ, ρ) ensemble with $\mathtt{h}^* < \mathtt{h}^{\mathrm{stab}}$. Also, consider the spatially-coupled LDPC (λ, ρ, N, w_0) ensemble with a fixed coupling window w_0, and a channel $\mathsf{c}(\mathtt{h})$ with $\mathtt{h} > \mathtt{h}^*$. Then, there exists an N_0 such that, for any $N > N_0$, the fixed point of density evolution resulting from $\underline{\Delta_0}$ initialization satisfies

$$\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\Delta_0}; \mathsf{c}(\mathtt{h})) \succ \underline{\Delta_\infty}.$$

Proof. First, choose $\mathtt{h} > \mathtt{h}^*$. Since $U_{\mathrm{s}}(\cdot\,; \mathsf{c}(\mathtt{h}))\colon \mathcal{X} \to \mathbb{R}$ is continuous and $\mathcal{X}$ is compact, $U_{\mathrm{s}}(\cdot\,; \mathsf{c}(\mathtt{h}))$ attains its minimum. Let a_* be a minimizer of $U_{\mathrm{s}}(\cdot\,; \mathsf{c}(\mathtt{h}))$. By Lemma 24, a_* is a fixed point of the single system DE. By assumption $\mathtt{h}^{\mathrm{stab}} > \mathtt{h}^*$, and $\mathtt{h} > \mathtt{h}^*$. Hence, by Lemma 30, $U_{\mathrm{s}}(\mathsf{a}_*; \mathsf{c}(\mathtt{h})) < 0$. Initialize the spatially-coupled LDPC (λ, ρ, N, w_0) system with $\underline{\mathsf{a}}_* = [\mathsf{a}_*, \ldots, \mathsf{a}_*]$. Since a_* is a fixed point of the

single system,

$$\begin{aligned}\mathsf{T}_{\mathrm{c}}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h}))_i &= \frac{1}{w}\sum_{k=0}^{w-1}\mathsf{c}(\mathtt{h})_{i-k} \circledast \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1}\rho^{\boxplus}(\mathsf{a}_*)\Big)\\ &\preceq \frac{1}{w}\sum_{k=0}^{w-1}\mathsf{c}(\mathtt{h}) \circledast \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1}\rho^{\boxplus}(\mathsf{a}_*)\Big)\\ &= \mathsf{c}(\mathtt{h}) \circledast \lambda^{\circledast}(\rho^{\boxplus}(\mathsf{a}_*)) = \mathsf{a}_*.\end{aligned}$$

That is, $\mathsf{T}_{\mathrm{c}}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h})) \preceq \underline{\mathsf{a}}_*$. Therefore, from the monotonicity of T_{c} by Lemma 34, $\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h}))$ exists and

$$\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h})) \preceq \mathsf{T}_{\mathrm{c}}^{(\ell+1)}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h})) \preceq \mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h})) \preceq \underline{\mathsf{a}}_*.$$

By Lemma 46 and the continuity of $U_{\mathrm{c}}(\cdot;\mathsf{c}(\mathtt{h}))$,

$$\begin{aligned}U_{\mathrm{c}}(\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h}));\mathsf{c}(\mathtt{h})) \le U_{\mathrm{c}}(\mathsf{T}_{\mathrm{c}}^{(\ell+1)}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h}));\mathsf{c}(\mathtt{h})) &\le U_{\mathrm{c}}(\mathsf{T}_{\mathrm{c}}^{(\ell)}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h}));\mathsf{c}(\mathtt{h}))\\ &\le U_{\mathrm{c}}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h})).\end{aligned}$$

Also, since all entries of $\underline{\mathsf{a}}_*$ are equal,

$$\begin{aligned}U_{\mathrm{c}}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h})) &= (2N+(w_0-1))U_{\mathrm{s}}(\mathsf{a}_*;\mathsf{c}(\mathtt{h})) + (w_0-1)\mathrm{H}\left(\mathsf{c}(\mathtt{h}) \circledast L^{\circledast}(\rho^{\boxplus}(\mathsf{a}_*))\right)\\ &\le (2N+(w_0-1))U_{\mathrm{s}}(\mathsf{a}_*;\mathsf{c}(\mathtt{h})) + w_0 - 1.\end{aligned}$$

Since $U_{\mathrm{s}}(\mathsf{a}_*;\mathsf{c}(\mathtt{h})) < 0$, we can choose large enough N_0 such that for all $N > N_0$, $U_{\mathrm{c}}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h})) < 0$. Therefore, $U_{\mathrm{c}}(\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h}));\mathsf{c}(\mathtt{h})) \le U_{\mathrm{c}}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h})) < 0$, and, since $U_{\mathrm{c}}(\underline{\Delta_\infty};\mathsf{c}(\mathtt{h})) = 0$, this implies that $\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h})) \neq \underline{\Delta_\infty}$. Since $\underline{\Delta_0} \succeq \underline{\mathsf{a}}_*$, $\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\Delta_0};\mathsf{c}(\mathtt{h})) \succeq \mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\mathsf{a}}_*;\mathsf{c}(\mathtt{h}))$. Hence, $\mathsf{T}_{\mathrm{c}}^{(\infty)}(\underline{\Delta_0};\mathsf{c}(\mathtt{h})) \succ \underline{\Delta_\infty}$. □

II.E LOW-DENSITY GENERATOR-MATRIX ENSEMBLES

II.E.1 Single System

Low-density generator-matrix (LDGM) ensembles are a class of linear codes that have a sparse generator-matrix representation. An example of a Tanner graph representation of an LDGM code is provided in Fig. II.4. The term LDGM(λ,ρ) denotes the LDGM ensemble with information-node degree distribution λ and generator-node degree distribution ρ from the edge perspective. An equivalent representation

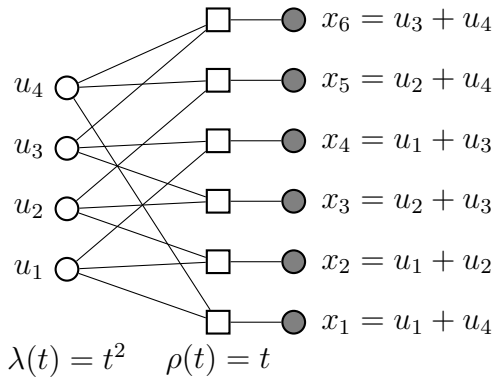

Figure II.4: The Tanner graph representation of an LDGM code with left-degree 3 and right-degree 2. The leftmost nodes u_i's are the information-nodes and the square nodes are generator-nodes. The rightmost nodes in gray represent the code-bits.

---

in terms of the node perspective degree distributions L, R is given by

$$\lambda(t) = \frac{L'(t)}{L'(1)}, \qquad \rho(t) = \frac{R'(t)}{R'(1)}.$$

LDGM codes are amenable to techniques similar to that of their counterpart, LDPC codes. However, a key issue here is that these codes have non-negligible error floors. One mathematical difficulty that arises from this is that the desired fixed point of DE is non-trivial and depends on the channel parameter. This poses a great challenge when characterizing thresholds, convergence, etc. Nevertheless, LDGM codes are an attractive option for rateless codes [45], [46], and in lossy source compression [47], [48]. See Section [38, Section 7.5] for an introduction to LDGM codes.

The analysis of LDGM codes, and their coupled variant, is very similar to that of the LDPC codes. Thus, we keep the same notation for analogous quantities.

The evolution of message distributions is characterized by the DE described by

$$\tilde{\mathsf{x}}^{(\ell+1)} = \lambda^{\circledast}(\mathsf{c} \boxtimes \rho^{\boxtimes}(\tilde{\mathsf{x}}^{(\ell)})), \tag{II.11}$$

where $\tilde{\mathsf{x}}^{(\ell)}$ denotes the message distribution at the output of information-nodes after ℓ iterations of message-passing, and c represents the channel LLR distribution. When the iterative system in (II.11) is initialized with a, the information-node output after ℓ iterations is denoted by $\mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a};\mathsf{c})$. The distribution after one iteration is therefore

$\mathsf{T}_{\mathrm{s}}^{(1)}(\mathsf{a};\mathsf{c})$, or shortly, $\mathsf{T}_{\mathrm{s}}(\mathsf{a};\mathsf{c})$. If the sequence of measures $\{\mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a};\mathsf{c})\}$ converges in $(\mathcal{X}, d_{\mathrm{H}})$, then its limit is denoted by $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c})$.

The DE update operator T_{s} satisfies exactly the same monotonicity properties as in Lemma 18. To avoid repetition, we do not state them explicitly.

We note that Δ_∞ is *not* a fixed point of (II.11), which is in stark contrast to LDPC codes. If this system is initialized with Δ_∞, then $\mathsf{T}_{\mathrm{s}}(\Delta_\infty;\mathsf{c}) \succeq \Delta_\infty$. As such, the sequence $\{\mathsf{T}_{\mathrm{s}}^{(\ell)}(\Delta_\infty;\mathsf{c})\}$ converges to the fixed point $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\Delta_\infty;\mathsf{c})$. If x is any fixed point of (II.11), since $\mathsf{x} \succeq \Delta_\infty$, by the monotonicity of T_{s},

$$\mathsf{x} = \mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{x};\mathsf{c}) \succeq \mathsf{T}_{\mathrm{s}}^{(\infty)}(\Delta_\infty;\mathsf{c}).$$

Thus, $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\Delta_\infty;\mathsf{c})$ is the *minimal* fixed point.

Definition 48: The *minimal fixed point* for the LDGM(λ,ρ) ensemble with channel c is defined to be

$$\mathsf{f}_0(\mathsf{c}) \triangleq \mathsf{T}_{\mathrm{s}}^{(\infty)}(\Delta_\infty;\mathsf{c}).$$

We also denote this by f_0 when the context is clear.

The following definition of the potential functional is essentially the negative of the trial-entropy or the replica-symmetric free entropy calculated in [39, Equation 6.2]. Also, in Section II.H.7.3, we briefly show the calculations to derive this potential from the Bethe formalism.

Definition 49: The potential functional $U_{\mathrm{s}}\colon \mathcal{X}\times\mathcal{X}\to\mathbb{R}$ for the LDGM(λ,ρ) ensemble with a channel c is defined as

$$\begin{aligned} U_{\mathrm{s}}(\mathsf{x};\mathsf{c}) = {} & \tfrac{L'(1)}{R'(1)}\mathrm{H}\left(\mathsf{c}\circledast R^{\circledast}(\mathsf{x})\right) - L'(1)\mathrm{H}\left(\mathsf{x}\circledast\mathsf{c}\circledast\rho^{\circledast}(\mathsf{x})\right) + L'(1)\mathrm{H}\left(\mathsf{c}\circledast\rho^{\circledast}(\mathsf{x})\right) \\ & - \mathrm{H}\left(L^{\circledast}(\mathsf{c}\circledast\rho^{\circledast}(\mathsf{x}))\right) - \tfrac{L'(1)}{R'(1)}\mathrm{H}\left(\mathsf{c}\right). \end{aligned}$$

The directional derivative of the potential functional gives rise to the DE update in (II.11). Using Proposition 5, we have the following result similar to Lemma 23.

Lemma 50: The directional derivative of the potential functional with respect to $\mathsf{x}\in\mathcal{X}$, in the direction $\mathsf{y}\in\mathcal{X}_{\mathrm{d}}$, is given by

$$\mathrm{d}_{\mathsf{x}}U_{\mathrm{s}}(\mathsf{x};\mathsf{c})[\mathsf{y}] = L'(1)\mathrm{H}\left(\left[\mathsf{T}_{\mathrm{s}}(\mathsf{x};\mathsf{c})-\mathsf{x}\right]\circledast\left[\mathsf{c}\circledast\rho'^{\circledast}(\mathsf{x})\circledast\mathsf{y}\right]\right).$$

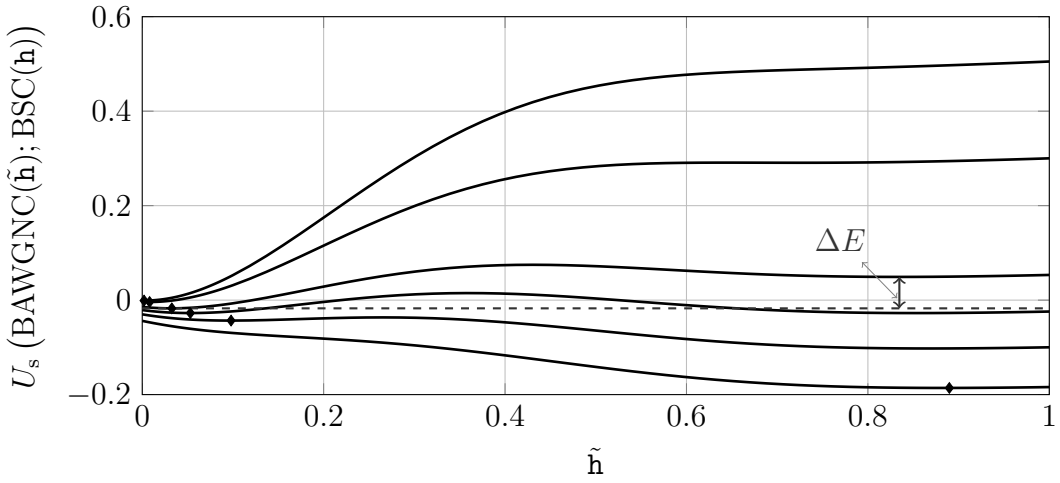

Figure II.5: Potential functional for an LDGM(λ,ρ) ensemble with $\lambda(t)=t^8$ and $\rho(t)=\frac{3}{50}+\frac{6}{50}t+\frac{9}{50}t^2+\frac{12}{50}t^3+\frac{20}{50}t^4$ over a binary symmetric channel with entropy $\mathtt{h}$. The values of $\mathtt{h}$ for these curves are, from the top to bottom, 0.37, 0.4529, 0.56, 0.5902, 0.62, 0.66. The other input to the potential functional is the binary AWGN channel (BAWGNC) with entropy $\tilde{\mathtt{h}}$. The choice of BAWGNC distribution for the first argument in $U_{\mathrm{s}}(\cdot\,;\cdot)$ is arbitrary. The marked points denote the minimal fixed points f_0.

Similar to Lemma 24, we can also show that the minimum of the potential functional for a fixed c occurs at a fixed point of the DE.

Definition 51: For the LDGM(λ,ρ) ensemble with a channel $\mathsf{c}\in\mathcal{X}$, define

i) The basin of attraction to $\mathsf{f}_0(\mathsf{c})$ as the set

$$\mathcal{V}(\mathsf{c})=\{\mathsf{x}\in\mathcal{X}\mid \mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{x};\mathsf{c})=\mathsf{f}_0(\mathsf{c})\}.$$

ii) The *energy gap* as

$$\Delta E(\mathsf{c})\triangleq\inf_{\mathsf{x}\in\mathcal{X}\setminus\mathcal{V}(\mathsf{c})}U_{\mathrm{s}}(\mathsf{x};\mathsf{c})-U_{\mathrm{s}}(\mathsf{f}_0(\mathsf{c});\mathsf{c}),$$

with the convention that the infimum over the empty set is ∞.

Fig. II.5 illustrates the potential functional of an LDGM ensemble over a BSC channel with

$$\lambda(t)=t^8,\qquad\qquad \rho(t)=\frac{3}{50}+\frac{6}{50}t+\frac{9}{50}t^2+\frac{12}{50}t^3+\frac{20}{50}t^4.$$

A few observations are in order. At small values of $\mathtt{h}$, the minimal fixed point $\mathsf{f}_0(\mathsf{c}(\mathtt{h}))$ determines the error floor of these ensembles. As we increase $\mathtt{h}$ beyond 0.4529, another fixed point appears in the right (from initializing DE with Δ_0), and this fixed point governs the DE performance. For $\mathtt{h} < 0.5902$, the energy gap $\Delta E(\mathsf{c}(\mathtt{h})) > 0$ stays positive. The range of $\mathtt{h}$ for which the energy gap stays positive is important, as this characterizes the performance of spatially-coupled codes. For large values of $\mathtt{h}$, the fixed point resulting from Δ_0 initialization and the minimal fixed point coincide. We emphasize that these observations are *only qualitative* as this two-dimensional illustration does not characterize the behavior of $U_{\mathrm{s}}(\cdot\,;\mathsf{c})$ over all $\mathcal{X}$.

By Definition 51(ii), $\Delta E(\mathsf{c}(\mathtt{h}))$ is a difference of two functions varying in $\mathtt{h}$. For general LDGM ensembles, whether the energy gap is monotone as a function of $\mathtt{h}$ is not known. This poses a difficulty when defining the potential threshold. We circumvent this by stating the threshold saturation theorem differently, and perhaps less elegantly, than LDPC ensembles. More precisely, the result we have for LDGM ensembles is the following (Theorem 61): If $\Delta E(\mathsf{c}) > 0$, then, for a large enough coupling window w, any DE fixed point of the spatially-coupled system is elementwise better (in the degradation order) than the *minimal fixed point* of the single system, $\mathsf{f}_0(\mathsf{c})$.

It is conjectured [39, Section X] that the region where $\Delta E(\mathsf{c}) > 0$ characterizes the MAP decoding performance. Accordingly, when $\Delta E(\mathsf{c}) > 0$, the potential functional is minimized at $\mathsf{f}_0(\mathsf{c})$ and therefore the value of $L^{\circledast}(\mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{f}_0(\mathsf{c})))$ under the error probability functional [38, Definition 4.53] characterizes the bit-error rate of the MAP decoder. Moreover, when $\Delta E(\mathsf{c}) < 0$, the MAP decoder performance is strictly worse than the one characterized by $L^{\circledast}(\mathsf{c}\boxtimes\rho^{\boxtimes}(\mathsf{f}_0(\mathsf{c})))$. Thus, if the conjecture in [39, Section X] is true, then the BP performance of the spatially-coupled ensemble and the MAP performance of the single system coincide.

II.E.2 Coupled System

The construction of spatially-coupled LDGM ensemble is similar to that of spatially coupled LDPC ensembles and we refer the reader to Section II.C.2 for an elaborate treatment. A performance analysis of spatially-coupled LDGM ensembles first appeared in [49]. The information-node groups are placed at positions in $\mathcal{N}_v = \{1, 2, \cdots, 2N\}$, and the generator-node groups at $\mathcal{N}_c = \{1, 2, \cdots, N_w\}$, where

$N_w = 2N + w - 1$. The DE update at generator-node inputs is given by

$$\mathsf{x}_i^{(\ell+1)} = \frac{1}{w}\sum_{k=0}^{w-1} \lambda^{\circledast}\left(\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{x}_{i-k+j}^{(\ell)}); \varepsilon_{i-k}\right), \tag{II.12}$$

for $i \in \mathcal{N}_c$, where $\mathsf{x}_i = \Delta_\infty$ when $i \notin \mathcal{N}_c$ and the shorthand $\lambda^{\circledast}(\mathsf{x}; \varepsilon_i)$ denotes

$$\lambda^{\circledast}(\mathsf{x}; \varepsilon_i) = \begin{cases} \lambda^{\circledast}(\mathsf{x}) & \text{if } i \in \mathcal{N}_v, \\ \Delta_\infty & \text{otherwise.} \end{cases}$$

We refer to the system characterized by (II.12) as the (λ, ρ, N, w) spatially-coupled LDGM ensemble.

A few of the terms that appear in the summation on the RHS of (II.12) will be Δ_∞ and these represent the boundary condition that gets decoding started. When the spatially-coupled LDGM system is initialized with $\underline{\mathsf{x}} = \underline{\Delta_0}$, the information at the boundary propagates inward and this induces a nondecreasing degradation ordering on positions $1, \ldots, \lceil N_w/2 \rceil$ and a nonincreasing degradation ordering on positions $\lceil N_w/2 \rceil + 1, \ldots, N_w$. This ordering results in a degraded maximum at position $i_0 = N + \lceil \frac{w-1}{2} \rceil$.

As seen in Section II.E.1, the minimal fixed point f_0 plays a crucial role in the performance of the LDGM ensembles under iterative decoding. Spatially-coupled LDGM ensembles are no exception. The minimal fixed point f_0 of the single system is also crucial for the spatially-coupled system. Changing the boundary in (II.12) from Δ_∞ to f_0 therefore facilitates the proof of threshold saturation for these ensembles.

Definition 52: The *modified system* is defined by the following update,

$$\mathsf{x}_i^{(\ell+1)} = \frac{1}{w}\sum_{k=0}^{w-1} \lambda^{\circledast}\left(\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{x}_{i-k+j}^{(\ell)}); \delta_{i-k}\right),$$

for $i \in \{1, \ldots, i_0\}$, and $\mathsf{x}_i^{(\ell+1)} = \mathsf{x}_{i_0}^{(\ell+1)}$ for $i_0 < i \leq N_w$, $\mathsf{x}_i = \mathsf{f}_0$ when $i \notin \mathcal{N}_c$. The shorthand $\lambda^{\circledast}(\mathsf{x}; \delta_i)$ represents

$$\lambda^{\circledast}(\mathsf{x}; \delta_i) = \begin{cases} \lambda^{\circledast}(\mathsf{x}) & \text{if } i \in \mathcal{N}_v, \\ \mathsf{f}_0 & \text{otherwise.} \end{cases}$$

In comparison to (II.12), the modified system here differs both in the boundary condition and the saturation constraint $\mathsf{x}_i = \mathsf{x}_{i_0}$ for $i_0 < i \leq N_w$. When the modified system and spatially-coupled system have the same initialization, as DE progresses, the distributions of the modified system will be degraded with respect to that of spatially-coupled system in (II.12). Again, the modified system serves as an upper bound to the spatially-coupled system. The DE updates for both spatially-coupled and modified system satisfy the monotonicity properties listed in Lemma 34. For brevity, we do not state them explicitly.

If the modified system is initialized with $\underline{\mathsf{x}}^{(0)} = \underline{\Delta_0}$, then $\underline{\mathsf{x}}^{(\ell+1)} \preceq \underline{\mathsf{x}}^{(\ell)}$ and $\underline{\mathsf{x}}^{(\ell)} \succeq \underline{\mathsf{f}_0}$ for all ℓ. To see this, suppose $\underline{\mathsf{x}}^{(\ell)} \succeq \underline{\mathsf{f}_0}$ for some ℓ (e.g., this is automatically true when $\ell = 0$). Observing the modified system DE update for $1 \leq i \leq i_0$,

$$\begin{aligned}
\mathsf{x}_i^{(\ell+1)} &= \frac{1}{w}\sum_{k=0}^{w-1} \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{x}_{i-k+j}^{(\ell)}); \delta_{i-k}\Big) \\
&\overset{(a)}{\succeq} \frac{1}{w}\sum_{k=0}^{w-1} \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{f}_0); \delta_{i-k}\Big) \\
&= \frac{1}{w}\sum_{k=0}^{w-1} \lambda^{\circledast}\Big(\mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{f}_0); \delta_{i-k}\Big) \\
&\overset{(b)}{=} \lambda^{\circledast}\left(\mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{f}_0)\right) \overset{(c)}{=} \mathsf{f}_0,
\end{aligned}$$

where (a) follows since $\underline{\mathsf{x}}^{(\ell)} \succeq \underline{\mathsf{f}_0}$, while (b) and (c) follow since f_0 is a fixed point of the single system DE. Thus, the sequence of measure vectors $\{\underline{\mathsf{x}}^{(\ell)}\}$ satisfies $\underline{\mathsf{x}}^{(\ell)} \succeq \underline{\mathsf{x}}^{(\ell+1)}$, $\underline{\mathsf{x}}^{(\ell)} \succeq \underline{\mathsf{f}_0}$, and consequently $\{\underline{\mathsf{x}}^{(\ell)}\}$ converges to a fixed point $\underline{\mathsf{x}}$ with $\underline{\mathsf{x}} \succeq \underline{\mathsf{f}_0}$. We also have the following result analogous to Lemma 36.

Lemma 53: The fixed point $\underline{\mathsf{x}}$ of the modified system resulting from $\underline{\Delta_0}$ initialization satisfies

$$\mathsf{x}_i \succeq \mathsf{x}_{i-1} \succeq \mathsf{f}_0, \quad 2 \leq i \leq N_w.$$

Below, we define the coupled potential for LDGM ensembles. Unlike LDPC codes, the coupled potential here and the properties that follow pertain exclusively to the modified system due to the difference in boundary conditions. The key difference in our proof strategy for LDGM codes is to tweak the coupled potential to reflect

the modified boundary and show that this modified potential still has the desired properties.

Definition 54: The coupled potential functional $U_\mathrm{c}\colon \mathcal{X}^{N_w} \times \mathcal{X} \to \mathbb{R}$ for a modified system is defined by

$$
\begin{aligned}
U_\mathrm{c}(\underline{\mathsf{x}};\mathsf{c}) &\triangleq \\
& L'(1)\sum_{i=1}^{N_w}\left[\frac{1}{R'(1)}\mathrm{H}\left(\mathsf{c}\boxast R^{\boxast}(\mathsf{x}_i)\right) - \frac{1}{R'(1)}\mathrm{H}\left(\mathsf{c}\right) - \mathrm{H}\left(\mathsf{x}_i \boxast \mathsf{c} \boxast \rho^{\boxast}(\mathsf{x}_i)\right)\right] \\
&+ L'(1)\sum_{i=1}^{N_w}\mathrm{H}\left(\mathsf{c}\boxast\rho^{\boxast}(\mathsf{x}_i)\right) - \sum_{i=1}^{2N}\mathrm{H}\left(L^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1}\mathsf{c}\boxast\rho^{\boxast}(\mathsf{x}_{i+j})\Big)\right) \\
&- L'(1)\sum_{i=1}^{w-1}\left[\frac{w-i}{w}\mathrm{H}\left(\mathsf{f}_0\circledast\left[\mathsf{c}\boxast\rho^{\boxast}(\mathsf{x}_i)\right]\right) + \frac{i}{w}\mathrm{H}\left(\mathsf{f}_0\circledast\left[\mathsf{c}\boxast\rho^{\boxast}(\mathsf{x}_{2N+i})\right]\right)\right].
\end{aligned}
\tag{II.13}
$$

The last two terms of (II.13) are not present in (II.6). These additional terms are necessary to reflect the modified boundary. Proofs of Lemmas 55, 56 are nearly identical to their analogues, Lemmas 38, 39, respectively.

Lemma 55: The directional derivative of the potential functional in (II.13) with respect to $\underline{\mathsf{x}} \in \mathcal{X}^{N_w}$, evaluated in the direction $\underline{\mathsf{y}} \in \mathcal{X}_\mathrm{d}^{N_w}$ is given by

$$
\begin{aligned}
&\mathrm{d}_{\underline{\mathsf{x}}}U_\mathrm{c}(\underline{\mathsf{x}};\mathsf{c})[\underline{\mathsf{y}}] = \\
&L'(1)\sum_{i=1}^{N_w}\mathrm{H}\left(\left[\frac{1}{w}\sum_{k=0}^{w-1}\lambda^{\circledast}\left(\frac{1}{w}\sum_{j=0}^{w-1}\mathsf{c}\boxast\rho^{\boxast}(\mathsf{x}_{i-k+j});\delta_{i-k}\right) - \mathsf{x}_i\right]\boxast\left[\mathsf{c}\boxast\rho'^{\boxast}(\mathsf{x}_i)\boxast\mathsf{y}_i\right]\right).
\end{aligned}
\tag{II.14}
$$

Lemma 56: The second-order directional derivative of the potential functional in

(II.13) with respect to $\underline{\mathsf{x}}$, evaluated in the direction $[\underline{\mathsf{y}}, \underline{\mathsf{z}}] \in \mathcal{X}_{\mathrm{d}}^{N_w} \times \mathcal{X}_{\mathrm{d}}^{N_w}$ is given by

$$\mathrm{d}^2_{\underline{\mathsf{x}}} U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c})[\underline{\mathsf{y}}, \underline{\mathsf{z}}] = \tag{II.15}$$

$$L'(1)\rho''(1) \sum_{i=1}^{N_w} \mathrm{H}\left(\left[\frac{1}{w} \sum_{k=0}^{w-1} \lambda^{\circledast}\left(\frac{1}{w} \sum_{j=0}^{w-1} \mathsf{c} \circledast \rho^{\circledast}(\mathsf{x}_{i-k+j}); \delta_{i-k} \right) \circledast \mathsf{c} \circledast \frac{\rho''^{\circledast}(\mathsf{x}_i)}{\rho''(1)} \right] \circledast \mathsf{y}_i \circledast \mathsf{z}_i \right)$$

$$- L'(1)\rho''(1) \sum_{i=1}^{N_w} \mathrm{H}\left(\left[\mathsf{x}_i \circledast \mathsf{c} \circledast \frac{\rho''^{\circledast}(\mathsf{x}_i)}{\rho''(1)} \right] \circledast \mathsf{y}_i \circledast \mathsf{z}_i \right)$$

$$- L'(1)\rho'(1) \sum_{i=1}^{N_w} \mathrm{H}\left(\mathsf{c} \circledast \frac{\rho'^{\circledast}(\mathsf{x}_i)}{\rho'(1)} \circledast \mathsf{y}_i \circledast \mathsf{z}_i \right) - \frac{L'(1)\lambda'(1)\rho'(1)^2}{w} \sum_{i=1}^{N_w} \sum_{m=\max\{i-(w-1),1\}}^{\min\{i+(w-1),N_w\}} \cdots$$

$$\mathrm{H}\left(\frac{1}{w} \sum_{k=0}^{w-1} \frac{\lambda'^{\circledast}\left(\frac{1}{w} \sum_{j=0}^{w-1} \mathsf{c} \circledast \rho^{\circledast}(\mathsf{x}_{i-k+j}); \delta_{i-k} \right)}{\lambda'(1)} \circledast \left[\mathsf{c} \circledast \frac{\rho'^{\circledast}(\mathsf{x}_i)}{\rho'(1)} \circledast \mathsf{y}_i \right] \circledast \left[\mathsf{c} \circledast \frac{\rho'^{\circledast}(\mathsf{x}_m)}{\rho'(1)} \circledast \mathsf{z}_m \right] \right),$$

where $\lambda'^{\circledast}(\mathsf{x}; \delta_i)$ denotes

$$\lambda'^{\circledast}(\mathsf{x}; \delta_i) = \begin{cases} \lambda'^{\circledast}(\mathsf{x}) & \text{if } i \in \mathcal{N}_v, \\ 0 & \text{otherwise.} \end{cases}$$

II.F THRESHOLD SATURATION FOR LDGM ENSEMBLES

The proof strategy for threshold saturation of spatially-coupled LDGM ensembles is similar to that of spatially-coupled LDPC ensembles. It is clear that f_0 plays a role similar to that of Δ_∞ for LDPC ensembles. The shift operator in Definition 57 is adjusted accordingly. Explicit characterization of the change in coupled potential due to shift is stated in Lemma 58. The proof for this lemma is considerably different from that of its counterpart in LDPC section, and it is detailed in Section II.H.5.1.

Lemmas 59 and 60 characterize the first- and second-order variations in the coupled potential at a non-trivial fixed point. Theorem 61 states the threshold saturation result. Proofs of Lemma 59, Lemma 60 and Theorem 61 are nearly identical to that of their counterparts in LDPC section, requiring only straightforward changes from Δ_∞ to f_0. We skip these proofs for brevity.

Definition 57: The shift operator $\mathsf{S} : \mathcal{X}^{N_w} \to \mathcal{X}^{N_w}$ is defined pointwise by

$$[\mathsf{S}(\underline{\mathsf{x}})]_1 \triangleq \mathsf{f}_0, \qquad [\mathsf{S}(\underline{\mathsf{x}})]_i \triangleq \mathsf{x}_{i-1}, \quad 2 \le i \le N_w.$$

Lemma 58: Let $\underline{\mathsf{x}} \in \mathcal{X}^{N_w}$ be such that $\underline{\mathsf{x}} \succeq \underline{\mathsf{f}_0} \triangleq [\mathsf{f}_0, \cdots, \mathsf{f}_0]$ and $\mathsf{x}_i = \mathsf{x}_{i_0}$, for $i_0 \leq i \leq N_w$. Also suppose $i_0 \leq 2N$. Then the change in the potential functional for a modified system associated with the shift operator is bounded by

$$U_{\mathrm{c}}(\mathsf{S}(\underline{\mathsf{x}}); \mathsf{c}) - U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c}) \leq U_{\mathrm{s}}(\mathsf{f}_0; \mathsf{c}) - U_{\mathrm{s}}(\mathsf{x}_{i_0}; \mathsf{c}) \tag{II.16}$$

Proof. See Section II.H.5.1. □

Lemma 59: If $\underline{\mathsf{x}} \succ \underline{\mathsf{f}_0}$ is a fixed point of the modified system resulting from $\underline{\Delta_0}$ initialization, then

$$\mathrm{d}_{\underline{\mathsf{x}}} U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c})[\mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}] = 0,$$

and moreover, x_{i_0} is not in the basin of attraction to f_0 (i.e., $\mathsf{x}_{i_0} \notin \mathcal{V}(\mathsf{c})$).

Lemma 58, Lemma 59, and Definition 51(ii) therefore imply that, for a non-trivial fixed point $\underline{\mathsf{x}}$ resulting from initializing the modified system with $\underline{\Delta_0}$,

$$U_{\mathrm{c}}(\mathsf{S}(\underline{\mathsf{x}}); \mathsf{c}) - U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c}) \leq U_{\mathrm{s}}(\mathsf{f}_0; \mathsf{c}) - U_{\mathrm{s}}(\mathsf{x}_{i_0}; \mathsf{c}) \leq -\Delta E(\mathsf{c}).$$

We note that while the shift bound in Lemma 58 requires $i_0 \leq 2N$, which is satisfied by choosing $N > \lceil \frac{w-1}{2} \rceil$, this restriction has no bearing on Theorem 61. This is because for a fixed w, distributions of spatially-coupled systems with larger N are degraded with respect to that of systems with smaller N.

Lemma 60: Suppose $\underline{\mathsf{x}}$ is a fixed point of the modified system resulting from $\underline{\Delta_0}$ initialization. Then

$$\left| \mathrm{d}^2_{\underline{\mathsf{x}}_1} U_{\mathrm{c}}(\underline{\mathsf{x}}_1; \mathsf{c})[\mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}, \mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}] \right| \leq \frac{K_{\lambda,\rho}}{w},$$

where the constant $K_{\lambda,\rho} \triangleq L'(1)\left(2\rho''(1) + \rho'(1) + 2\lambda'(1)\rho'(1)^2\right)$ is independent of N and w.

Theorem 61: Fix the LDGM(λ, ρ) ensemble and a BMS channel c with $\Delta E(\mathsf{c}) > 0$. For the (λ, ρ, N, w) spatially-coupled LDGM ensemble with $w > K_{\lambda,\rho}/(2\Delta E(\mathsf{c}))$, any

fixed point $\underline{\mathsf{x}}$ of density evolution satisfies

$$\mathsf{x}_i \preceq \mathsf{f}_0(\mathsf{c}), \quad 1 \leq i \leq N_w.$$

II.G CONCLUSIONS

In this chapter, a proof of threshold saturation, based on potential functions, is provided for spatially-coupled codes over BMS channels. In particular, we show that for spatially-coupled irregular LDPC codes over a BMS channel, the belief-propagation decoding threshold saturates to the conjectured MAP threshold. For LDGM codes, although the notion of thresholds is not systematically defined, a similar result holds. A converse to the threshold saturation result is also provided for LDPC codes. This result reiterates the generality of the threshold saturation phenomenon, which is now evident from many observations and proofs that span a wide variety of systems.

The approach taken here can be seen as analyzing the average Bethe free entropy in the large-system limit. We also believe that this approach can be extended to more general graphical models by computing their average Bethe free entropy.

II.H APPENDIX

II.H.1 A Metric Topology on $\mathcal{X}$

This section establishes a metric topology on $\mathcal{X}$ that is homeomorphic to the weak topology on the set of probability measures on $[0, 1]$. The given metric is closely related to the entropy functional. The reader is assumed to be familiar with the notation in Section II.B.

For $\mathsf{x} \in \mathcal{X}$, recall from Proposition 7,

$$\mathrm{H}(\mathsf{x}) = 1 - \sum_{k=1}^{\infty} \gamma_k M_k(\mathsf{x}), \quad \text{where } \gamma_k = \frac{(\log 2)^{-1}}{2k(2k-1)}.$$

The entropy distance $d_{\mathrm{H}} \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is defined as

$$d_{\mathrm{H}}(\mathsf{x}_1, \mathsf{x}_2) \triangleq \sum_{k=1}^{\infty} \gamma_k \left| M_k(\mathsf{x}_1) - M_k(\mathsf{x}_2) \right|.$$

Endow the space of extended real numbers $\overline{\mathbb{R}} = [-\infty, \infty]$ with the metric given

by $d_{\overline{\mathbb{R}}}(\alpha_1, \alpha_2) = |\tanh(\alpha_1) - \tanh(\alpha_2)|$. Under this metric, $\overline{\mathbb{R}}$ is *compact*. We begin by establishing a bijection between the set of symmetric probability measures on $\overline{\mathbb{R}}$, $\mathcal{X}$, and the set of probability measures on $[0,1]$, denoted by $\mathbb{P}([0,1])$. This bijection is useful when characterizing the properties of the entropy distance d_{H}.

Remark: The role of the entropy distance d_{H} is similar to that of the Wasserstein metric in [19, Section II-H]. In fact, one could easily define a weighted Wasserstein metric where, like d_{H}, the distance between x_1 and x_2 is equal to $\mathrm{H}(\mathsf{x}_1 - \mathsf{x}_2)$ if $\mathsf{x}_1 \succeq \mathsf{x}_2$. The relationship between such a weighted Wasserstein metric and d_{H} warrants further attention.

The function defined by $\psi\colon [-\infty, \infty] \to [0,1]$, $\psi(\alpha) = \tanh^2(\frac{\alpha}{2})$ is continuous. Consider the pushforward measure from $\mathcal{X}$ to $\mathbb{P}([0,1])$ induced by ψ,

$$\begin{aligned} \Psi\colon \mathcal{X} &\to \mathbb{P}([0,1]) \\ \mathsf{x} &\mapsto \hat{\mathsf{x}}, \end{aligned}$$

where $\hat{\mathsf{x}}(A) = \mathsf{x}(\psi^{-1}(A))$ for all Borel sets $A \in \mathcal{B}([0,1])$. Below, for any $\mathsf{x} \in \mathcal{X}$, we denote $\hat{\mathsf{x}}$ for $\Psi(\mathsf{x})$. For any measurable $f\colon [0,1] \to \mathbb{R}$, $\int f d\hat{\mathsf{x}} = \int (f \circ \psi) d\mathsf{x}$. This immediately implies that $\int \alpha^k \hat{\mathsf{x}}(d\alpha) = \int \tanh^{2k}\left(\frac{\alpha}{2}\right) \mathsf{x}(d\alpha)$. Thus, k-th moments of $\hat{\mathsf{x}}$ are given by $M_k(\mathsf{x})$.

Lemma: The function $\Psi\colon \mathcal{X} \to \mathbb{P}([0,1])$ defined above is a bijection.

Proof. For injectivity of Ψ, consider $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}$ such that $\hat{\mathsf{x}}_1 = \hat{\mathsf{x}}_2$. Clearly, $\mathsf{x}_1(\{0\}) = \mathsf{x}_2(\{0\})$. Suppose E is a Borel set in $\mathcal{B}((0,\infty])$ and $A_E = \psi(E)$. We have

$$\mathsf{x}_1(\psi^{-1}(A_E)) = \mathsf{x}_2(\psi^{-1}(A_E)),$$

which implies

$$\begin{aligned} \int_{-E} \mathsf{x}_1(d\alpha) + \int_E \mathsf{x}_1(d\alpha) &= \int_{-E} \mathsf{x}_2(d\alpha) + \int_E \mathsf{x}_2(d\alpha), \\ \int_E (1 + e^{-\alpha}) \mathsf{x}_1(d\alpha) &= \int_E (1 + e^{-\alpha}) \mathsf{x}_2(d\alpha), \end{aligned}$$

due to symmetry. Since $1 + e^{-\alpha}$ is non-zero, $\mathsf{x}_1(E) = \mathsf{x}_2(E)$ for all $E \in \mathcal{B}((0,\infty])$.

Again by symmetry,

$$\mathsf{x}_1(-E) = \int_E e^{-\alpha} \mathsf{x}_1(d\alpha) = \int_E e^{-\alpha} \mathsf{x}_2(d\alpha) = \mathsf{x}_2(-E).$$

This implies that $\mathsf{x}_1(E) = \mathsf{x}_2(E)$ for all $E \in \mathcal{B}(\overline{\mathbb{R}})$, and consequently, $\mathsf{x}_1 = \mathsf{x}_2$. Hence, Ψ is injective.

For surjectivity, suppose $\mu \in \mathbb{P}([0,1])$. Define measures $\mathsf{x}_1, \mathsf{x}_2$ on $[0,\infty]$ such that for $E \in \mathcal{B}([0,\infty])$,

$$\mathsf{x}_1(E) = \mu(\psi(E)), \qquad\qquad \mathsf{x}_2(E) = \int_E \frac{1}{1+e^{-\alpha}} \mathsf{x}_1(d\alpha).$$

Extend x_2 to $[-\infty,\infty]$ by defining x as

$$\begin{aligned} \mathsf{x}(E) &= \mathsf{x}_2(E), \quad \text{for } E \in \mathcal{B}((0,\infty]), \\ \mathsf{x}(\{0\}) &= 2\mathsf{x}_2(\{0\}), \\ \mathsf{x}(E) &= \int_{-E} e^{-\alpha} \mathsf{x}_2(d\alpha), \quad \text{for } E \in \mathcal{B}([-\infty,0)). \end{aligned}$$

Then, x is a symmetric probability measure on $[-\infty,\infty]$, and $\hat{\mathsf{x}} = \mu$. Hence Ψ is surjective. □

Proposition: The set of symmetric probability measures with the entropy distance $(\mathcal{X}, d_{\mathrm{H}})$ is a metric space.

Proof. It is easy to see that $d_{\mathrm{H}}(\cdot,\cdot)$ is non-negative, symmetric, and satisfies the triangle inequality. For d_{H} to be a metric, it suffices to show that $d_{\mathrm{H}}(\mathsf{x}_1, \mathsf{x}_2) = 0$ implies $\mathsf{x}_1 = \mathsf{x}_2$. Let $d_{\mathrm{H}}(\mathsf{x}_1, \mathsf{x}_2) = 0$. Note that $d_{\mathrm{H}}(\mathsf{x}_1, \mathsf{x}_2) = 0$ iff $M_k(\mathsf{x}_1) = M_k(\mathsf{x}_2)$ for all $k \in \mathbb{N}$. Thus $\int \alpha^k \hat{\mathsf{x}}_1(d\alpha) = \int \alpha^k \hat{\mathsf{x}}_2(d\alpha)$, for all $k \in \mathbb{N}$. By the Hausdorff moment problem [50, Theorem VII.3.1], $\hat{\mathsf{x}}_1 = \hat{\mathsf{x}}_2$. By injectivity of Ψ, $\mathsf{x}_1 = \mathsf{x}_2$. Thus d_{H} is a metric on $\mathcal{X}$. □

Proposition: The metric topology $(\mathcal{X}, d_{\mathrm{H}})$ is homeomorphic to the weak topology on $\mathbb{P}([0,1])$.

Proof. It suffices to show that Ψ and Ψ^{-1} are continuous. Suppose $\mu_n \to \mu$ weakly in $\mathbb{P}([0,1])$. Since $x^k \colon [0,1] \to [0,1]$ is a bounded continuous function for $k \in \mathbb{N}$,

$\int \alpha^k \mu_n(d\alpha) \to \int \alpha^k \mu(d\alpha)$. But this implies $M_k(\Psi^{-1}(\mu_n)) \to M_k(\Psi^{-1}(\mu))$. Hence Ψ^{-1} is continuous.

For the continuity of Ψ, let $\mathsf{x}_n \xrightarrow{d_{\mathrm{H}}} \mathsf{x}$ in $\mathcal{X}$. That is $\int \alpha^k \hat{\mathsf{x}}_n(d\alpha) \to \int \alpha^k \hat{\mathsf{x}}(d\alpha)$, and consequently, $\int p(\alpha)\hat{\mathsf{x}}_n(d\alpha) \to \int p(\alpha)\hat{\mathsf{x}}(d\alpha)$, for any polynomial $p\colon [0,1] \to \mathbb{R}$. By an application of the Stone-Weirstrass theorem [51, Theorem 4.45], polynomials are dense in the set of continuous functions on $[0,1]$ under the supremum norm, $C[0,1]$. This implies $\int f(\alpha)\hat{\mathsf{x}}_n(d\alpha) \to \int f(\alpha)\hat{\mathsf{x}}(d\alpha)$, for any $f \in C([0,1])$. Thus $\hat{\mathsf{x}}_n \to \hat{\mathsf{x}}$ weakly, and this establishes the continuity of Ψ. □

Corollary: The metric topology $(\mathcal{X}, d_{\mathrm{H}})$ is compact and separable. Since compact metric spaces are complete, it is also a Polish space.

Proposition: The functionals $\mathrm{H}\colon \mathcal{X} \to \mathbb{R}$ and $M_k\colon \mathcal{X} \to \mathbb{R}$ are continuous.

Proof. The continuity of H follows since $|\mathrm{H}(\mathsf{x}_1) - \mathrm{H}(\mathsf{x}_2)| \leq d_{\mathrm{H}}(\mathsf{x}_1, \mathsf{x}_2)$, while the continuity of $M_k(\cdot)$ follows from $|M_k(\mathsf{x}_1) - M_k(\mathsf{x}_2)| \leq \frac{1}{\gamma_k} d_{\mathrm{H}}(\mathsf{x}_1, \mathsf{x}_2)$. □

Proposition: If we endow $\mathcal{X} \times \mathcal{X}$ with the product topology, then the operators $\circledast\colon \mathcal{X} \times \mathcal{X} \to \mathcal{X}$ and $\circledcirc\colon \mathcal{X} \times \mathcal{X} \to \mathcal{X}$ are continuous.

Proof. Suppose $\mathsf{x}_{n,1} \xrightarrow{d_{\mathrm{H}}} \mathsf{x}_1$ and $\mathsf{x}_{n,2} \xrightarrow{d_{\mathrm{H}}} \mathsf{x}_2$. Below, we will show that $\mathsf{x}_{n,1} \circledast \mathsf{x}_{n,2} \xrightarrow{d_{\mathrm{H}}} \mathsf{x}_1 \circledast \mathsf{x}_2$ and $\mathsf{x}_{n,1} \circledcirc \mathsf{x}_{n,2} \xrightarrow{d_{\mathrm{H}}} \mathsf{x}_1 \circledcirc \mathsf{x}_2$. First, consider the operator $\circledast$.

$$\begin{aligned} d_{\mathrm{H}}(\mathsf{x}_{n,1} \circledast \mathsf{x}_{n,2}, \mathsf{x}_1 \circledast \mathsf{x}_2) &= \sum_{k=1}^{\infty} \gamma_k \left| M_k(\mathsf{x}_{n,1}) M_k(\mathsf{x}_{n,2}) - M_k(\mathsf{x}_1) M_k(\mathsf{x}_2) \right| \\ &\leq \sum_{k=1}^{\infty} \gamma_k \left| M_k(\mathsf{x}_{n,1}) - M_k(\mathsf{x}_1) \right| M_k(\mathsf{x}_{n,2}) \\ &\qquad + \sum_{k=1}^{\infty} \gamma_k \left| M_k(\mathsf{x}_{n,2}) - M_k(\mathsf{x}_2) \right| M_k(\mathsf{x}_1) \\ &\leq d_{\mathrm{H}}(\mathsf{x}_{n,1}, \mathsf{x}_1) + d_{\mathrm{H}}(\mathsf{x}_{n,2}, \mathsf{x}_2) \to 0. \end{aligned}$$

Thus $\circledast$ is continuous. For the operator $\circledcirc$, note that $\hat{\mathsf{x}}_{n,1} \to \hat{\mathsf{x}}_1$ weakly and $\hat{\mathsf{x}}_{n,2} \to \hat{\mathsf{x}}_2$ weakly. Let $\mu_n = \Psi(\mathsf{x}_{n,1} \circledcirc \mathsf{x}_{n,2})$. We have

$$\begin{aligned} M_k(\mathsf{x}_{n,1} \circledcirc \mathsf{x}_{n,2}) &= \int \tanh^{2k}\left(\tfrac{\alpha}{2}\right) (\mathsf{x}_{n,1} \circledcirc \mathsf{x}_{n,2})(d\alpha) \\ &= \int \alpha^k \mu_n(d\alpha) \end{aligned}$$

$$= \iint f_{\circledast,k}(\alpha_1,\alpha_2)\hat{\mathsf{x}}_{n,1}(d\alpha_1)\hat{\mathsf{x}}_{n,2}(d\alpha_2),$$

where the kernel $f_{\circledast,k}\colon [0,1]\times[0,1]\to\mathbb{R}$ is the continuous function given by

$$\begin{aligned} f_{\circledast,k}(\alpha_1,\alpha_2) = {} & \tfrac{1+\sqrt{\alpha_1\alpha_2}}{2}\tanh^{2k}\left(\tanh^{-1}(\sqrt{\alpha_1})+\tanh^{-1}(\sqrt{\alpha_2})\right) \\ & + \tfrac{1-\sqrt{\alpha_1\alpha_2}}{2}\tanh^{2k}\left(\tanh^{-1}(\sqrt{\alpha_1})-\tanh^{-1}(\sqrt{\alpha_2})\right). \end{aligned}$$

Since $f_{\circledast,k}$ is continuous and $\{\hat{\mathsf{x}}_{n,1}\}$, $\{\hat{\mathsf{x}}_{n,2}\}$ converge weakly,

$$\begin{aligned} \iint f_{\circledast,k}(\alpha_1,\alpha_2)\hat{\mathsf{x}}_{n,1}(d\alpha_1)\hat{\mathsf{x}}_{n,2}(d\alpha_2) &\to \iint f_{\circledast,k}(\alpha_1,\alpha_2)\hat{\mathsf{x}}_1(d\alpha_1)\hat{\mathsf{x}}_2(d\alpha_2) \\ &= \int \alpha^k \mu(d\alpha) = M_k(\mathsf{x}_1\circledast\mathsf{x}_2), \end{aligned}$$

where $\mu = \Psi(\mathsf{x}_1\circledast\mathsf{x}_2)$. Thus $M_k(\mathsf{x}_{n,1}\circledast\mathsf{x}_{n,2})\to M_k(\mathsf{x}_1\circledast\mathsf{x}_2)$, and consequently, $\mathsf{x}_{n,1}\circledast\mathsf{x}_{n,2}\xrightarrow{d_{\mathrm{H}}}\mathsf{x}_1\circledast\mathsf{x}_2$. This establishes the continuity of $\circledast$. □

Proposition: If a sequence of measures $\{\mathsf{x}_n\}_{n=1}^{\infty}$ satisfies $\mathsf{x}_{n+1}\succeq\mathsf{x}_n$ (respectively, $\mathsf{x}_{n+1}\preceq\mathsf{x}_n$), then $\mathsf{x}_n\xrightarrow{d_{\mathrm{H}}}\mathsf{x}$, for some $\mathsf{x}\in\mathcal{X}$ which satisfies $\mathsf{x}\succeq\mathsf{x}_n$ (respectively, $\mathsf{x}\preceq\mathsf{x}_n$) for all n.

Proof. We suppose $\mathsf{x}_{n+1}\succeq\mathsf{x}_n$ for $n\in\mathbb{N}$; the case where $\mathsf{x}_{n+1}\preceq\mathsf{x}_n$ follows similarly. Since the entropy functional preserves the order by degradation, $\mathrm{H}(\mathsf{x}_{n+1})\geq\mathrm{H}(\mathsf{x}_n)$. Since $0\leq\mathrm{H}(\mathsf{x})\leq 1$ for $\mathsf{x}\in\mathcal{X}$, $\{\mathrm{H}(\mathsf{x}_n)\}$ is a Cauchy sequence. For any $m>n$, since $\mathsf{x}_m\succeq\mathsf{x}_n$, $d_{\mathrm{H}}(\mathsf{x}_m,\mathsf{x}_n)=\mathrm{H}(\mathsf{x}_m)-\mathrm{H}(\mathsf{x}_n)\to 0$ as $m,n\to\infty$. Thus, the sequence $\{\mathsf{x}_n\}$ is Cauchy and as $(\mathcal{X},d_{\mathrm{H}})$ is complete, $\mathsf{x}_n\xrightarrow{d_{\mathrm{H}}}\mathsf{x}$ for some $\mathsf{x}\in\mathcal{X}$.

To show $\mathsf{x}\succeq\mathsf{x}_n$, in view of Definition 2, let f be a concave non-increasing function on $[0,1]$. Then, necessarily, f is continuous on $[0,1)$. First suppose f is continuous on $[0,1]$. We discuss the case where $f(1)<\lim_{\alpha\to 1}f(\alpha)$ separately. Since $\mathsf{x}_{n+1}\succeq\mathsf{x}_n$, for any $m>n$, $\mathsf{x}_m\succeq\mathsf{x}_n$. This implies

$$\begin{aligned} \int f\left(\left|\tanh\left(\tfrac{\alpha}{2}\right)\right|\right)\mathsf{x}_m(d\alpha) &\geq \int f\left(\left|\tanh\left(\tfrac{\alpha}{2}\right)\right|\right)\mathsf{x}_n(d\alpha), \\ \int (f\circ\sqrt{\cdot})d\hat{\mathsf{x}}_m &\geq \int (f\circ\sqrt{\cdot})d\hat{\mathsf{x}}_n, \\ \lim_{m\to\infty}\int (f\circ\sqrt{\cdot})d\hat{\mathsf{x}}_m &\geq \int (f\circ\sqrt{\cdot})d\hat{\mathsf{x}}_n, \end{aligned}$$

and, since $\hat{\mathsf{x}}_m \to \hat{\mathsf{x}}$ weakly and $f \circ \sqrt{\cdot}$ is continuous on $[0,1]$, $\lim_{m\to\infty} \int (f \circ \sqrt{\cdot}) d\hat{\mathsf{x}}_m = \int (f \circ \sqrt{\cdot}) d\hat{\mathsf{x}}$. Thus,

$$\int (f \circ \sqrt{\cdot}) d\hat{\mathsf{x}} \geq \int (f \circ \sqrt{\cdot}) d\hat{\mathsf{x}}_n,$$
$$\int f\left(\left|\tanh\left(\tfrac{\alpha}{2}\right)\right|\right) \mathsf{x}(d\alpha) \geq \int f\left(\left|\tanh\left(\tfrac{\alpha}{2}\right)\right|\right) \mathsf{x}_n(d\alpha).$$

Now suppose f is a concave, non-increasing function on $[0,1]$, but discontinuous at 1. Since f is bounded, to show $\int (f \circ \sqrt{\cdot}) d\hat{\mathsf{x}} \geq \int (f \circ \sqrt{\cdot}) d\hat{\mathsf{x}}_n$, we can assume f is non-negative by adding a suitable constant. Also, there exists a sequence of functions $\{f_m\}_{m=1}^{\infty}$ that are non-negative, non-increasing, continuous, concave and $f_m \leq f_{m+1}$, $f_m \to f$ pointwise. By the monotone convergence theorem [51, Theorem 2.14],

$$\int (f \circ \sqrt{\cdot}) d\hat{\mathsf{x}} = \lim_{m\to\infty} \int (f_m \circ \sqrt{\cdot}) d\hat{\mathsf{x}}, \qquad \int (f \circ \sqrt{\cdot}) d\hat{\mathsf{x}}_n = \lim_{m\to\infty} \int (f_m \circ \sqrt{\cdot}) d\hat{\mathsf{x}}_n.$$

Since f_m is continuous, from the arguments above, $\int (f_m \circ \sqrt{\cdot}) d\hat{\mathsf{x}} \geq \int (f_m \circ \sqrt{\cdot}) d\hat{\mathsf{x}}_n$. Consequently, $\int (f \circ \sqrt{\cdot}) d\hat{\mathsf{x}} \geq \int (f \circ \sqrt{\cdot}) d\hat{\mathsf{x}}_n$. Hence $\mathsf{x} \succeq \mathsf{x}_n$ for any n. □

We state the following result without proof as it is similar to the previous proposition.

Proposition: If $\{\mathsf{x}'_n\}_{n=1}^{\infty}$, $\{\mathsf{x}_n\}_{n=1}^{\infty}$ satisfy $\mathsf{x}'_n \succeq \mathsf{x}_n$ and $\mathsf{x}'_n \xrightarrow{d_{\mathrm{H}}} \mathsf{x}'$, $\mathsf{x}_n \xrightarrow{d_{\mathrm{H}}} \mathsf{x}$, then $\mathsf{x}' \succeq \mathsf{x}$.

II.H.2 Proofs from Section II.B

II.H.2.1 Proof of Proposition 1

By symmetry and since $f(0) = 0$ for an odd function,

$$\begin{aligned}
\int f(\alpha) \mathsf{x}(d\alpha) &= f(0)\mathsf{x}(\{0\}) + \int_{(0,\infty]} \left[f(\alpha) + f(-\alpha)e^{-\alpha}\right] \mathsf{x}(d\alpha) \\
&= \int_{(0,\infty]} f(\alpha)(1 - e^{-\alpha}) \, \mathsf{x}(d\alpha) \\
&= \int_{(0,\infty]} f(\alpha) \tanh\left(\tfrac{\alpha}{2}\right) (1 + e^{-\alpha}) \, \mathsf{x}(d\alpha) \\
&= \int f(\alpha) \tanh\left(\tfrac{\alpha}{2}\right) \, \mathsf{x}(d\alpha).
\end{aligned}$$

II.H.2.2 Proof of Proposition 6

i) Follows from $0 \leq \tanh^{2k}(\alpha) \leq 1$.

ii) Note that $f(\alpha) = -\alpha^{2k}$ is a concave decreasing function over $[0, 1]$. Since $\mathsf{x}_1 \succeq \mathsf{x}_2$, Definition 2 implies that

$$-M_k(\mathsf{x}_1) = I_f(\mathsf{x}_1) \geq I_f(\mathsf{x}_2) = -M_k(\mathsf{x}_2).$$

Thus, $M_k(\mathsf{x}_1) \leq M_k(\mathsf{x}_2)$.

iii) By the equivalent characterization of the operator $\boxtimes$,

$$\begin{aligned} M_k(\mathsf{x}_1 \boxtimes \mathsf{x}_2) &= \int \tanh^{2k}(\tfrac{\alpha}{2})(\mathsf{x}_1 \boxtimes \mathsf{x}_2)(d\alpha) \\ &\overset{(a)}{=} \iint \tanh^{2k}\Big(\tfrac{\tau^{-1}(\tau(\alpha_1)\tau(\alpha_2))}{2}\Big)\mathsf{x}_1(d\alpha_1)\mathsf{x}_2(d\alpha_2) \\ &= \iint \tanh^{2k}\big(\tfrac{\alpha_1}{2}\big)\tanh^{2k}\big(\tfrac{\alpha_2}{2}\big)\mathsf{x}_1(d\alpha_1)\mathsf{x}_2(d\alpha_2) \\ &= M_k(\mathsf{x}_1)M_k(\mathsf{x}_2), \end{aligned}$$

where $\tau(\alpha) = \tanh(\frac{\alpha}{2})$ in the RHS of (a).

iv) If $\mathsf{x} = \Delta_\infty$ (respectively, $\mathsf{x} = \Delta_0$), then it is easy to see that $M_k(\mathsf{x}) = 1$ (respectively, $M_k(\mathsf{x}) = 0$) for all k. The other direction follows from

$$0 < \tanh^{2k}(\alpha) \quad \text{if } \alpha \neq 0, \qquad\qquad 1 > \tanh^{2k}(\alpha) \quad \text{if } \alpha \neq \pm\infty,$$

and since the symmetry of the measure implies $\mathsf{x}(\{-\infty\}) = e^{-\infty}\mathsf{x}(\{\infty\}) = 0$.

II.H.2.3 Proof of Proposition 8

i) Using Proposition 7 and $(\mathsf{y}_1 \boxtimes \mathsf{y}_2)(\overline{\mathbb{R}}) = 0$ when $\mathsf{y}_1, \mathsf{y}_2 \in \mathcal{X}_\mathrm{d}$, we have the result.

ii) With the observation $\mathrm{H}(\mathsf{y}_1 \circledast \mathsf{y}_2) = -\mathrm{H}(\mathsf{y}_1 \boxtimes \mathsf{y}_2)$ from Proposition 5, the inequalities are trivial. It remains to show that $\mathsf{y} = 0$ when $\mathrm{H}(\mathsf{y} \boxtimes \mathsf{y}) = 0$. For this, let $\mathsf{y} = \mathsf{x}_1 - \mathsf{x}_2$ with $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}$, and observe that

$$\mathrm{H}(\mathsf{y} \boxtimes \mathsf{y}) = 0 \iff M_k(\mathsf{x}_1) = M_k(\mathsf{x}_2) \quad \text{for all } k.$$

The fact that $M_k(\mathsf{x}_1) = M_k(\mathsf{x}_2)$ for all k iff $\mathsf{x}_1 = \mathsf{x}_2$ follows as a consequence of the metric properties of the entropy functional; see Definition 10 and Proposition 11.

iii) Using the first part of this proposition and the inequalities $M_k(\mathsf{x}_1') \leq M_k(\mathsf{x}_1)$ and $M_k(\mathsf{x}_2') \leq M_k(\mathsf{x}_2)$, we have the result.

iv) Assume $\mathsf{x}_1 \succ \mathsf{x}_2$ and consider $\mathsf{x}_3 \neq \Delta_\infty$. To show $\mathrm{H}(\mathsf{x}_1 \circledast \mathsf{x}_3) > \mathrm{H}(\mathsf{x}_2 \circledast \mathsf{x}_3)$, observe that

$$\begin{aligned}\mathrm{H}(\mathsf{x}_1 \circledast \mathsf{x}_3) - \mathrm{H}(\mathsf{x}_2 \circledast \mathsf{x}_3) &= \mathrm{H}([\mathsf{x}_1 - \mathsf{x}_2] \circledast [\mathsf{x}_3 - \Delta_\infty]) \\ &= -\mathrm{H}([\mathsf{x}_1 - \mathsf{x}_2] \circledcirc [\mathsf{x}_3 - \Delta_\infty]) \quad \text{(Proposition 5)} \\ &= \sum_{k=0}^{\infty} \gamma_k [M_k(\mathsf{x}_2) - M_k(\mathsf{x}_1)][1 - M_k(\mathsf{x}_3)] > 0.\end{aligned}$$

The last inequality follows since $M_k(\mathsf{x}_3) < 1$ for all $k \in \mathbb{N}$ (from Proposition 6(iv)) and $M_k(\mathsf{x}_2) > M_k(\mathsf{x}_1)$ for some $k \in \mathbb{N}$ (see the proof of part ii of this proposition).

Now, consider $\mathsf{x}_3 \neq \Delta_0$. Again, we observe that

$$\mathrm{H}(\mathsf{x}_1 \circledcirc \mathsf{x}_3) - \mathrm{H}(\mathsf{x}_2 \circledcirc \mathsf{x}_3) = \mathrm{H}([\mathsf{x}_1 - \mathsf{x}_2] \circledcirc \mathsf{x}_3) = \sum_{k=0}^{\infty} \gamma_k [M_k(\mathsf{x}_2) - M_k(\mathsf{x}_1)] M_k(\mathsf{x}_3) > 0,$$

where the last inequality follows since $M_k(\mathsf{x}_3) > 0$ for all k and $M_k(\mathsf{x}_2) > M_k(\mathsf{x}_1)$ for some k.

II.H.2.4 Proof of Proposition 12

From [38, Problems 4.60-61], $2\mathfrak{E}(\mathsf{x}) \leq \mathrm{H}(\mathsf{x}) \leq \mathfrak{B}(\mathsf{x})$, where $\mathfrak{E}(\cdot)$ is the error functional $\mathfrak{E}(\mathsf{x}) \triangleq \frac{1}{2} \int e^{-(\alpha + |\alpha|)/2} \mathsf{x}(d\alpha)$. From [38, Lemma 4.66], for $n \geq 2$,

$$\frac{\alpha \mathfrak{B}(\mathsf{x})^{3/2}}{\sqrt{n}} \mathfrak{B}(\mathsf{x})^n \leq 2\mathfrak{E}(\mathsf{x}^{\circledast n}) \leq \mathfrak{B}(\mathsf{x})^n,$$

for a constant $\alpha > 0$. The above relations, together with $\mathfrak{B}(\mathsf{x}^{\circledast n}) = \mathfrak{B}(\mathsf{x})^n$, imply that $\lim_{n \to \infty} \frac{1}{n} \log \mathrm{H}(\mathsf{x}^{\circledast n}) = \log \mathfrak{B}(\mathsf{x})$.

II.H.3 Proofs From Section II.C

II.H.3.1 Proof of Lemma 24

The first statement follows from Lemma 23.

For the second part, suppose x is not a fixed point of single system DE. We discuss the cases $\mathsf{x} \neq \Delta_0$ and $\mathsf{x} = \Delta_0$ separately. First, consider $\mathsf{x} \neq \Delta_0$. The derivative in Lemma 23 in the direction $\mathsf{T}_{\mathrm{s}}(\mathsf{x};\mathsf{c}) - \mathsf{x}$ is

$$\mathrm{d}_{\mathsf{x}} U_{\mathrm{s}}(\mathsf{x};\mathsf{c})[\mathsf{T}_{\mathrm{s}}(\mathsf{x};\mathsf{c}) - \mathsf{x}] = L'(1)\mathrm{H}\Big((\mathsf{T}_{\mathrm{s}}(\mathsf{x};\mathsf{c}) - \mathsf{x})^{\circledast 2} \circledast \rho'^{\circledast}(\mathsf{x}) \Big).$$

From Proposition 8(ii), the above equation is strictly negative if $\mathsf{x} \neq \mathsf{T}_{\mathrm{s}}(\mathsf{x};\mathsf{c})$ and $\mathsf{x} \neq \Delta_0$. Thus, if $\mathsf{x} \neq \mathsf{T}_{\mathrm{s}}(\mathsf{x};\mathsf{c})$ and $\mathsf{x} \neq \Delta_0$,

$$\mathrm{d}_{\mathsf{x}} U_{\mathrm{s}}(\mathsf{x};\mathsf{c})[\mathsf{T}_{\mathrm{s}}(\mathsf{x};\mathsf{c}) - \mathsf{x}] < 0.$$

By definition, $\lim_{\delta \to 0} \frac{U_{\mathrm{s}}(\mathsf{x} + \delta[\mathsf{T}_{\mathrm{s}}(\mathsf{x};\mathsf{c}) - \mathsf{x}];\mathsf{c}) - U_{\mathrm{s}}(\mathsf{x};\mathsf{c})}{\delta} < 0$. Thus, there exists a $t \in (0,1]$ such that $U_{\mathrm{s}}\left(\mathsf{x} + t\left[\mathsf{T}_{\mathrm{s}}(\mathsf{x};\mathsf{c}) - \mathsf{x}\right];\mathsf{c}\right) < U_{\mathrm{s}}(\mathsf{x};\mathsf{c})$. Therefore, $U_{\mathrm{s}}(\mathsf{x};\mathsf{c})$ cannot be a minimum if x is not a fixed point and $\mathsf{x} \neq \Delta_0$.

Now, we consider the case $\mathsf{x} = \Delta_0$. Since x is not a fixed point, $\mathsf{T}_{\mathrm{s}}(\Delta_0;\mathsf{c}) \prec \Delta_0$. For notational convenience, let

$$\mathsf{x}_t = \mathsf{T}_{\mathrm{s}}(\Delta_0;\mathsf{c}) + t[\Delta_0 - \mathsf{T}_{\mathrm{s}}(\Delta_0;\mathsf{c})] \quad \text{for } t \in [0,1].$$

This implies for $t \in (0,1)$, $\mathsf{x}_0 \prec \mathsf{x}_t \prec \Delta_0$, and by the monotonicity of the operator T_{s}, $\mathsf{T}_{\mathrm{s}}(\mathsf{x}_t;\mathsf{c}) \preceq \mathsf{T}_{\mathrm{s}}(\Delta_0;\mathsf{c}) = \mathsf{x}_0 \prec \mathsf{x}_t$. Define $\phi\colon [0,1] \to \mathbb{R}$, $\phi(t) = U_{\mathrm{s}}(\mathsf{x}_t;\mathsf{c})$. As in Proposition 16, for $t \in (0,1)$,

$$\begin{aligned}
\phi'(t) &= \mathrm{d}_{\mathsf{x}_t} U_{\mathrm{s}}(\mathsf{x}_t;\mathsf{c})[\Delta_0 - \mathsf{x}_0] \\
&= -L'(1)\mathrm{H}\left([\mathsf{x}_t - \mathsf{T}_{\mathrm{s}}(\mathsf{x}_t;\mathsf{c})] \circledast [\Delta_0 - \mathsf{x}_0] \circledast \rho'^{\circledast}(\mathsf{x}_t)\right) \\
&= L'(1)\mathrm{H}\left([\mathsf{x}_t - \mathsf{T}_{\mathrm{s}}(\mathsf{x}_t;\mathsf{c})] \circledast \mathsf{x}_0 \circledast \rho'^{\circledast}(\mathsf{x}_t)\right) > 0,
\end{aligned}$$

by Proposition 3(ii), since $\mathsf{x}_t \succ \mathsf{T}_{\mathrm{s}}(\mathsf{x}_t;\mathsf{c})$, $\mathsf{x}_0 \neq \Delta_0$ and $\rho'^{\circledast}(\mathsf{x}_t) \neq \Delta_0$. Thus,

$$U_{\mathrm{s}}(\Delta_0;\mathsf{c}) = \phi(1) > \phi(0) = U_{\mathrm{s}}(\mathsf{x}_0;\mathsf{c}).$$

As such, $U_{\mathrm{s}}(\Delta_0;\mathsf{c})$ cannot be a minimum of $U_{\mathrm{s}}(\cdot\,;\mathsf{c})$.

Hence, the minimum of $U_{\mathrm{s}}(\cdot\,;\mathsf{c})$ can only occur at a density evolution fixed point.

II.H.3.2 Proof of Lemma 26

i) By Proposition 8(iv), $\mathrm{H}\left(\mathsf{c}_1 \circledast \mathsf{x}\right) > \mathrm{H}\left(\mathsf{c}_2 \circledast \mathsf{x}\right)$ if $\mathsf{x} \neq \Delta_\infty$. Thus, $U_{\mathrm{s}}(\mathsf{x};\mathsf{c}_1) < U_{\mathrm{s}}(\mathsf{x};\mathsf{c}_2)$ if $\mathsf{x} \neq \Delta_\infty$.

ii) Using monotonicity of the DE operator, $\mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a};\mathsf{c}_1) \succeq \mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a};\mathsf{c}_2)$. Thus, if $\mathsf{a} \in \mathcal{V}(\mathsf{c}_1)$, then $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c}_1) = \Delta_\infty$. Then, it is easy to show that $\mathsf{T}_{\mathrm{s}}^{(\ell)}(\mathsf{a};\mathsf{c}_2) \xrightarrow{d_{\mathrm{H}}} \Delta_\infty$. Thus $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{a};\mathsf{c}_2) = \Delta_\infty$, and $\mathsf{a} \in \mathcal{V}(\mathsf{c}_2)$.

iii) Follows from parts i and ii.

II.H.3.3 Proof of Lemma 29

i) If $\mathtt{h}^{\mathrm{stab}} = 1$, then the result is trivial; therefore we assume $\mathtt{h}^{\mathrm{stab}} < 1$. Consider any $\mathtt{h} > \mathtt{h}^{\mathrm{stab}}$. From [38, Section 4.9.2], $\mathcal{V}(\mathsf{c}(\mathtt{h})) = \{\Delta_\infty\}$, and by the continuity of $U_{\mathrm{s}}(\cdot\,;\mathsf{c}(\mathtt{h}))$ at Δ_∞, $\Delta E(\mathsf{c}(\mathtt{h})) \leq 0$. This implies $\mathtt{h} \geq \mathtt{h}^*$ by Definition 28(iii). Thus $\mathtt{h}^* \leq \mathtt{h}^{\mathrm{stab}}$.

ii) If $\mathtt{h} < \mathtt{h}^{\mathrm{stab}}$, there exists an $\varepsilon > 0$ such that for all x with $\mathrm{H}\left(\mathsf{x}\right) < \varepsilon$, $\mathsf{x} \in \mathcal{V}(\mathsf{c}(\mathtt{h}))$ [38, Section 4.9.2]. Thus, if $d_{\mathrm{H}}(\mathsf{x}, \Delta_\infty) < \varepsilon$, then $d_{\mathrm{H}}(\Delta_\infty, \mathsf{x}) = \mathrm{H}\left(\mathsf{x}\right) < \varepsilon$, and hence $\mathsf{x} \in \mathcal{V}(\mathsf{c}(\mathtt{h}))$. Thus, there is an ε-ball around Δ_∞ which is in $\mathcal{V}(\mathsf{c}(\mathtt{h}))$.

II.H.3.4 Proof of Lemma 30

If $\mathtt{h}^* = 1$, then the statement of the lemma is vacuous; suppose $\mathtt{h}^* < 1$. Let $\mathtt{h} > \mathtt{h}^*$. By assumption, $\mathtt{h}^* < \mathtt{h}^{\mathrm{stab}}$, and thus there exists $\mathtt{h}' < \mathtt{h}$ such that $\mathtt{h}^* < \mathtt{h}' < \mathtt{h}^{\mathrm{stab}}$. Since $\mathtt{h}' < \mathtt{h}^{\mathrm{stab}}$, by Lemma 29,

$$\Delta_\infty \in (\mathcal{V}(\mathsf{c}(\mathtt{h}')))^o \qquad \Longrightarrow \qquad \Delta_\infty \notin \overline{\mathcal{X} \backslash \mathcal{V}(\mathsf{c}(\mathtt{h}'))}.$$

Moreover, $\overline{\mathcal{X} \backslash \mathcal{V}(\mathsf{c}(\mathtt{h}'))}$ is compact and $U_{\mathrm{s}}(\cdot\,;\mathsf{c}(\mathtt{h}'))$ is continuous. Therefore, the infimum $\inf_{\mathsf{x} \in \overline{\mathcal{X} \backslash \mathcal{V}(\mathsf{c}(\mathtt{h}'))}} U_{\mathrm{s}}(\mathsf{x};\mathsf{c}(\mathtt{h}'))$ is achieved at some $\mathsf{a} \neq \Delta_\infty$. By Lemma 26(i),

$U_{\mathrm{s}}(\mathsf{a};\mathsf{c}(\mathtt{h}))$ is strictly decreasing in $\mathtt{h}$. Therefore,

$$\begin{aligned}\min_{\mathsf{x}\in\mathcal{X}} U_{\mathrm{s}}(\mathsf{x};\mathsf{c}(\mathtt{h})) &\le U_{\mathrm{s}}(\mathsf{a};\mathsf{c}(\mathtt{h}))\\ &< U_{\mathrm{s}}(\mathsf{a};\mathsf{c}(\mathtt{h}')) \quad (\text{Since } \mathtt{h}'<\mathtt{h})\\ &= \inf_{\mathsf{x}\in\mathcal{X}\setminus\mathcal{V}(\mathsf{c}(\mathtt{h}'))} U_{\mathrm{s}}(\mathsf{x};\mathsf{c}(\mathtt{h}'))\\ &\le \inf_{\mathsf{x}\in\mathcal{X}\setminus\mathcal{V}(\mathsf{c}(\mathtt{h}'))} U_{\mathrm{s}}(\mathsf{x};\mathsf{c}(\mathtt{h}'))\\ &= \Delta E(\mathsf{c}(\mathtt{h}'))\le 0 \quad (\text{Since } \mathtt{h}'>\mathtt{h}^*).\end{aligned}$$

Hence, $\min_{\mathsf{x}\in\mathcal{X}} U_{\mathrm{s}}(\mathsf{x};\mathsf{c}(\mathtt{h}))<0$, and there exists an $\mathsf{x}\in\mathcal{X}$ such that $U_{\mathrm{s}}(\mathsf{x};\mathsf{c}(\mathtt{h}))<0$.

II.H.3.5 Proof of Lemma 36

Since the modified system is initialized with $\underline{\mathsf{x}}^{(0)}=\underline{\Delta_0}$, $\mathsf{x}_i^{(0)}\succeq\mathsf{x}_{i-1}^{(0)}$. Suppose at some iteration ℓ, $\mathsf{x}_i^{(\ell)}\succeq\mathsf{x}_{i-1}^{(\ell)}$. If $i>i_0$, then due to the saturation constraint in the modified system, $\mathsf{x}_i^{(\ell+1)}=\mathsf{x}_{i_0}^{(\ell+1)}$, $\mathsf{x}_i^{(\ell+1)}\succeq\mathsf{x}_{i-1}^{(\ell+1)}$. For $1\le i\le i_0$, by observing (II.5),

$$\mathsf{x}_i^{(\ell+1)}-\mathsf{x}_{i-1}^{(\ell+1)}=\frac{1}{w}\mathsf{c}_i\circledast\lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1}\rho^{\boxplus}(\mathsf{x}_{i+j}^{(\ell)})\Big)-\frac{1}{w}\mathsf{c}_{i-w}\circledast\lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1}\rho^{\boxplus}(\mathsf{x}_{i-w+j}^{(\ell)})\Big).$$

Note that $\mathsf{c}_i=\mathsf{c}$ if $i\in\mathcal{N}_v$ and $\mathsf{c}_i=\Delta_\infty$ otherwise. At this point, we need to consider two cases: 1) $2N\ge i_0$ 2) $2N<i_0$.

When $2N\ge i_0$, for any $1\le i\le i_0$, $i\in\mathcal{N}_v$, which implies $\mathsf{c}_i=\mathsf{c}$ and $\mathsf{c}_i\succeq\mathsf{c}_{i-w}$. Since $\mathsf{x}_i^{(\ell)}\succeq\mathsf{x}_{i-1}^{(\ell)}$, we see that $\mathsf{x}_i^{(\ell+1)}\succeq\mathsf{x}_{i-1}^{(\ell+1)}$.

When $2N<i_0$, for $2N<i\le i_0$, we note that $\mathsf{c}_i=\Delta_\infty$. However, $2N<i_0=N+\lfloor\frac{w}{2}\rfloor$ implies $N<\lfloor\frac{w}{2}\rfloor$. Thus, if $2N<i\le i_0$, then we have

$$2N-w<i-w\le i_0-w=N+\lfloor\tfrac{w}{2}\rfloor-w\le N-\lfloor\tfrac{w}{2}\rfloor<0.$$

As such, $\mathsf{c}_{i-w}=\Delta_\infty$. Here again, $\mathsf{c}_i\succeq\mathsf{c}_{i-w}$ and $\mathsf{x}_i^{(\ell+1)}\succeq\mathsf{x}_{i-1}^{(\ell+1)}$.

By letting $\ell\to\infty$, we have $\mathsf{x}_i\succeq\mathsf{x}_{i-1}$ by Proposition 11, where $\underline{\mathsf{x}}$ is the limit of $\{\underline{\mathsf{x}}^{(\ell)}\}$.

II.H.3.6 Proof of Lemma 38

The linearity of the entropy functional and the properties of the operators $\circledast$ and $\boxtimes$ (e.g., see Proposition 14) allow one to write

$$\mathrm{d}_{\underline{\mathsf{x}}} U_{\mathrm{c}}(\underline{\mathsf{x}};\mathsf{c})[\underline{\mathsf{y}}] = \sum_{i=1}^{N_w} \mathrm{d}_{\mathsf{x}_i} U_{\mathrm{c}}(\underline{\mathsf{x}};\mathsf{c})[\mathsf{y}_i].$$

As in the proof of Lemma 23, using the duality rule for entropy for differences of symmetric measures, the derivatives of the first three terms of U_{c} in (II.6) are

$$\begin{aligned}
\mathrm{d}_{\mathsf{x}_i} \mathrm{H}\left(R^{\circledast}(\mathsf{x}_i)\right)[\mathsf{y}_i] &= R'(1)\mathrm{H}\left(\rho^{\circledast}(\mathsf{x}_i) \circledast \mathsf{y}_i\right), \\
\mathrm{d}_{\mathsf{x}_i} \mathrm{H}\left(\rho^{\circledast}(\mathsf{x}_i)\right)[\mathsf{y}_i] &= \mathrm{H}\left(\rho'^{\circledast}(\mathsf{x}_i) \circledast \mathsf{y}_i\right), \\
\mathrm{d}_{\mathsf{x}_i} \mathrm{H}\left(\mathsf{x}_i \circledast \rho^{\circledast}(\mathsf{x}_i)\right)[\mathsf{y}_i] &= \mathrm{H}\left(\rho^{\circledast}(\mathsf{x}_i) \circledast \mathsf{y}_i\right) + \mathrm{H}\left(\rho'^{\circledast}(\mathsf{x}_i) \circledast \mathsf{y}_i\right) \\
&\quad - \mathrm{H}\left(\mathsf{x}_i \circledast \left[\rho'^{\circledast}(\mathsf{x}_i) \circledast \mathsf{y}_i\right]\right).
\end{aligned}$$

For the final term in (II.6), observe that if $w \leq i \leq 2N$, since there are exactly w components containing x_i, its derivative with respect to x_i is

$$\tfrac{L'(1)}{w} \sum_{k=0}^{w-1} \mathrm{H}\Big(\mathsf{c} \circledast \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \rho^{\boxtimes}(\mathsf{x}_{i-k+j})\Big) \circledast (\rho'^{\boxtimes}(\mathsf{x}_i) \boxtimes \mathsf{y}_i)\Big).$$

If $1 \leq i < w$, derivative of the final term in (II.6) with respect to x_i is

$$\tfrac{L'(1)}{w} \sum_{k=0}^{i-1} \mathrm{H}\Big(\mathsf{c} \circledast \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \rho^{\boxtimes}(\mathsf{x}_{i-k+j})\Big) \circledast (\rho'^{\boxtimes}(\mathsf{x}_i) \boxtimes \mathsf{y}_i)\Big).$$

This can be written as

$$\tfrac{L'(1)}{w} \sum_{k=0}^{w-1} \mathrm{H}\Big(\mathsf{c}_{i-k} \circledast \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \rho^{\boxtimes}(\mathsf{x}_{i-k+j})\Big) \circledast (\rho'^{\boxtimes}(\mathsf{x}_i) \boxtimes \mathsf{y}_i)\Big),$$

where $\mathsf{c}_i = \mathsf{c}$ when $1 \leq i \leq 2N$ and $\mathsf{c}_i = \Delta_\infty$ otherwise. This is because $\mathrm{H}\left(\Delta_\infty \circledast \mathsf{x}\right) = 0$ for any x, and hence the additional terms that are added evaluate to zero. A similar expression holds when $2N < i \leq N_w$. Combining these observations, the derivative

of the final term in (II.6) with respect to x_i for $1 \le i \le N_w$ is

$$\tfrac{L'(1)}{w} \sum_{k=0}^{w-1} \mathrm{H}\Big(\mathsf{c}_{i-k} \circledast \lambda^{\circledast}\Big(\frac{1}{w} \sum_{j=0}^{w-1} \rho^{\boxtimes}(\mathsf{x}_{i-k+j}) \Big) \circledast (\rho'^{\boxtimes}(\mathsf{x}_i) \boxtimes \mathsf{y}_i) \Big),$$

which is $L'(1)\mathrm{H}\left(\mathsf{T}_{\mathrm{c}}(\underline{\mathsf{x}};\mathsf{c})_i \circledast (\rho'^{\boxtimes}(\mathsf{x}_i) \boxtimes \mathsf{y}_i)\right)$. Consolidating these four terms and using Proposition 5 results in (II.7).

II.H.3.7 Proof of Lemma 39

We have

$$\mathrm{d}^2_{\underline{\mathsf{x}}} U_{\mathrm{c}}(\underline{\mathsf{x}};\mathsf{c})[\underline{\mathsf{y}},\underline{\mathsf{z}}] = \sum_{m=1}^{N_w} \sum_{i=1}^{N_w} \mathrm{d}_{\mathsf{x}_m} \left(\mathrm{d}_{\mathsf{x}_i} U_{\mathrm{c}}(\underline{\mathsf{x}};\mathsf{c})[\mathsf{y}_i] \right)[\mathsf{z}_m].$$

Using the calculations for $\mathrm{d}_{\mathsf{x}_i} U_{\mathrm{c}}(\underline{\mathsf{x}};\mathsf{c})[\mathsf{y}_i]$ in Section II.H.3.6, it is tedious but straightforward to obtain the desired result.

II.H.4 Proofs From Section II.D

II.H.4.1 Proof of Lemma 41

Due to the boundary condition $\mathsf{x}_i = \mathsf{x}_{i_0}$, for $i_0 \le i \le N_w$, the only terms that contribute to $U_{\mathrm{c}}(\mathsf{S}(\underline{\mathsf{x}});\mathsf{c}) - U_{\mathrm{c}}(\underline{\mathsf{x}};\mathsf{c})$ are given by

$$\begin{aligned} U_{\mathrm{c}}(\mathsf{S}(\underline{\mathsf{x}});\mathsf{c}) - U_{\mathrm{c}}(\underline{\mathsf{x}};\mathsf{c}) = &- \tfrac{L'(1)}{R'(1)} \mathrm{H}\left(R^{\boxtimes}(\mathsf{x}_{N_w})\right) - L'(1)\mathrm{H}\left(\rho^{\boxtimes}(\mathsf{x}_{N_w})\right) \\ &+ L'(1)\mathrm{H}\left(\mathsf{x}_{N_w} \boxtimes \rho^{\boxtimes}(\mathsf{x}_{N_w})\right) + \mathrm{H}\Big(\mathsf{c} \circledast L^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \rho^{\boxtimes}(\mathsf{x}_{2N+j})\Big)\Big) \\ &- \mathrm{H}\Big(\mathsf{c} \circledast L^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \rho^{\boxtimes}(\mathsf{x}_j)\Big)\Big), \quad \text{where } \mathsf{x}_0 = \Delta_\infty. \end{aligned}$$

Since $\mathsf{x}_{2N+j} \preceq \mathsf{x}_{N_w} = \mathsf{x}_{i_0}$ for $0 \le j \le w-1$ and the contribution from the last term is negative,

$$\begin{aligned} U_{\mathrm{c}}(\mathsf{S}(\underline{\mathsf{x}});\mathsf{c}) - U_{\mathrm{c}}(\underline{\mathsf{x}};\mathsf{c}) &\le -\tfrac{L'(1)}{R'(1)} \mathrm{H}\left(R^{\boxtimes}(\mathsf{x}_{N_w})\right) - L'(1)\mathrm{H}\left(\rho^{\boxtimes}(\mathsf{x}_{N_w})\right) \\ &\quad + L'(1)\mathrm{H}\left(\mathsf{x}_{N_w} \boxtimes \rho^{\boxtimes}(\mathsf{x}_{N_w})\right) + \mathrm{H}\Big(\mathsf{c} \circledast L^{\circledast}\Big(\rho^{\boxtimes}(\mathsf{x}_{N_w})\Big)\Big) \\ &= -U_{\mathrm{s}}(\mathsf{x}_{N_w};\mathsf{c}) = -U_{\mathrm{s}}(\mathsf{x}_{i_0};\mathsf{c}). \end{aligned}$$

II.H.4.2 Proof of Lemma 42

Since $\underline{\mathsf{x}}$ is a fixed point of the modified system, $\mathsf{x}_i = \mathsf{T}_{\mathrm{c}}(\underline{\mathsf{x}};\mathsf{c})_i$, for $1 \le i \le i_0$. Since $\mathsf{x}_i = \mathsf{x}_{i-1}$ for $i_0 < i \le N_w$, we have $[\mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}]_i = 0$. The first result follows from applying these relations to the directional derivative given in Lemma 38.

Below, we show that $\mathsf{x}_{i_0} \notin \mathcal{V}(\mathsf{c})$. By assumption, we know that $\underline{\mathsf{x}} \succ \underline{\Delta_\infty}$, and by Lemma 36, $\mathsf{x}_i \succeq \mathsf{x}_{i-1}$. Thus, $\mathsf{x}_{i_0} \succ \Delta_\infty$. Also,

$$\begin{aligned}
\mathsf{x}_{i_0} = \mathsf{T}_{\mathrm{c}}(\underline{\mathsf{x}};\mathsf{c})_{i_0} &= \frac{1}{w}\sum_{k=0}^{w-1} \mathsf{c}_{i_0-k} \circledast \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \rho^{\boxtimes}(\mathsf{x}_{i_0-k+j})\Big) \\
&\preceq \frac{1}{w}\sum_{k=0}^{w-1} \mathsf{c}_{i_0-k} \circledast \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \rho^{\boxtimes}(\mathsf{x}_{i_0})\Big) \\
&\preceq \mathsf{c} \circledast \lambda^{\circledast}\left(\rho^{\boxtimes}(\mathsf{x}_{i_0})\right) \\
&= \mathsf{T}_{\mathrm{s}}(\mathsf{x}_{i_0};\mathsf{c}).
\end{aligned}$$

Hence, by Lemma 18, $\mathsf{T}_{\mathrm{s}}^{(\infty)}(\mathsf{x}_{i_0};\mathsf{c}) \succeq \mathsf{T}_{\mathrm{s}}(\mathsf{x}_{i_0};\mathsf{c}) \succeq \mathsf{x}_{i_0} \succ \Delta_\infty$. Thus $\mathsf{x}_{i_0} \notin \mathcal{V}(\mathsf{c})$.

II.H.4.3 Proof of Lemma 43

Let $\underline{\mathsf{y}} = \mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}$, with componentwise decomposition

$$\mathsf{y}_i = [\mathsf{S}(\underline{\mathsf{x}}) - \underline{\mathsf{x}}]_i = \mathsf{x}_{i-1} - \mathsf{x}_i,$$

where $\mathsf{x}_i = \Delta_\infty$ for $i < 1$. Since $\underline{\mathsf{x}}$ is a fixed point of the modified system, if $i > i_0$, due to the saturation constraint, $\mathsf{x}_i = \mathsf{x}_{i-1}$. If $1 \le i \le i_0$, then using the update in (II.5) gives

$$\mathsf{x}_{i-1} - \mathsf{x}_i = \frac{1}{w}\mathsf{c}_{i-w} \circledast \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \rho^{\boxtimes}(\mathsf{x}_{i-w+j})\Big) - \frac{1}{w}\mathsf{c}_i \circledast \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \rho^{\boxtimes}(\mathsf{x}_{i+j})\Big).$$

Thus, $\mathsf{y}_i = \mathsf{x}_{i-1} - \mathsf{x}_i$ is of the form $\frac{1}{w}\mathsf{a}_i - \frac{1}{w}\mathsf{b}_i$, $\mathsf{a}_i, \mathsf{b}_i \in \mathcal{X}$ for all i (if $i > i_0$, $\mathsf{a}_i = \mathsf{b}_i$). From Lemma 39 and (II.8), the first three terms of the second-order directional derivative are of the form, for some $\mathsf{d} \in \mathcal{X}$,

$$\mathrm{H}\left(\mathsf{d} \circledast \mathsf{y}_i \circledast \mathsf{y}_i\right) = \frac{1}{w}\mathrm{H}\left(\mathsf{d}_3 \circledast (\mathsf{b}_i - \mathsf{a}_i) \circledast (\mathsf{x}_i - \mathsf{x}_{i-1})\right),$$

by linearity of the entropy functional. From Lemma 36, $\mathsf{x}_i \succeq \mathsf{x}_{i-1}$, and by Proposition 9, this term is absolutely bounded by

$$|\mathrm{H}\left(\mathsf{d} \circledast \mathsf{y}_i \circledast \mathsf{y}_i\right)| \leq \frac{1}{w}\mathrm{H}\left(\mathsf{x}_i - \mathsf{x}_{i-1}\right).$$

The final term is of the form, for some $\mathsf{d}_1, \mathsf{d}_2, \mathsf{d}_3, \mathsf{d}_4, \mathsf{d}_5 \in \mathcal{X}$,

$$\begin{aligned}
|\mathrm{H}\left(\mathsf{d}_1 \circledast [\mathsf{d}_2 \circledast \mathsf{y}_m] \circledast [\mathsf{d}_3 \circledast \mathsf{y}_i]\right)| &= |\mathrm{H}\left([\mathsf{d}_1 \circledast (\mathsf{d}_2 \circledast \mathsf{y}_m)] \circledast [\mathsf{d}_3 \circledast \mathsf{y}_i]\right)| \quad \text{(Proposition 5)} \\
&= |\mathrm{H}\left(\mathsf{d}_3 \circledast [\mathsf{d}_1 \circledast (\mathsf{d}_2 \circledast \mathsf{y}_m)] \circledast \mathsf{y}_i\right)| \\
&= \frac{1}{w}\left|\mathrm{H}\left(\mathsf{d}_3 \circledast [\mathsf{d}_5 - \mathsf{d}_4] \circledast [\mathsf{x}_i - \mathsf{x}_{i-1}]\right)\right| \ (\mathsf{y}_m = \tfrac{1}{w}\mathsf{a}_m - \tfrac{1}{w}\mathsf{b}_m) \\
&\leq \frac{1}{w}\mathrm{H}\left(\mathsf{x}_i - \mathsf{x}_{i-1}\right). \quad \text{(Proposition 9)}
\end{aligned}$$

By telescoping, one observes $\sum_{i=1}^{N_w} \mathrm{H}\left(\mathsf{x}_i - \mathsf{x}_{i-1}\right) = \mathrm{H}\left(\mathsf{x}_{N_w} - \Delta_\infty\right) \leq 1$. Combining these observations, the triangle inequality provides

$$\begin{aligned}
\left|\mathrm{d}^2_{\underline{\mathsf{x}}_1} U_{\mathrm{c}}(\underline{\mathsf{x}}_1; \mathsf{c})[\underline{\mathsf{y}}, \underline{\mathsf{y}}]\right| &\leq L'(1)\left(2\rho''(1)\frac{1}{w} + \rho'(1)\frac{1}{w} + 2w\frac{\lambda'(1)\rho'(1)^2}{w}\frac{1}{w}\right) \\
&= \frac{L'(1)\left(2\rho''(1) + \rho'(1) + 2\lambda'(1)\rho'(1)^2\right)}{w}.
\end{aligned}$$

II.H.5 Proofs from Section II.F

II.H.5.1 Proof of Lemma 58

Due to the boundary condition $\mathsf{x}_i = \mathsf{x}_{i_0}$ for $i_0 < i \leq N_w$ and by assumption $i_0 \leq 2N$, the terms that contribute to $U_{\mathrm{c}}(\mathsf{S}(\underline{\mathsf{x}}); \mathsf{c}) - U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c})$ are given by

$$\begin{aligned}
U_{\mathrm{c}}(\mathsf{S}(\underline{\mathsf{x}}); \mathsf{c}) - U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c}) = U_{\mathrm{s}}(\mathsf{f}_0; \mathsf{c}) - U_{\mathrm{s}}(\mathsf{x}_{i_0}; \mathsf{c}) &+ L'(1)\mathrm{H}\Big(\mathsf{f}_0 \circledast \Big[\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \circledast \rho^{\circledast}(\mathsf{x}_j)\Big]\Big) \\
&- L'(1)\mathrm{H}\left(\mathsf{f}_0 \circledast \left[\mathsf{c} \circledast \rho^{\circledast}(\mathsf{f}_0)\right]\right) - \mathrm{H}\Big(L^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \circledast \rho^{\circledast}(\mathsf{x}_j)\Big)\Big) \\
&+ \mathrm{H}\left(L^{\circledast}(\mathsf{c} \circledast \rho^{\circledast}(\mathsf{f}_0))\right),
\end{aligned}$$

where $\mathsf{x}_0 = \mathsf{f}_0$. It suffices to show that the contribution from the last four terms is negative. Define $F\colon \mathcal{X}^w \to \mathbb{R}$ by

$$F(\underline{\mathsf{x}}) = L'(1)\mathrm{H}\Big(\mathsf{f}_0 \circledast \Big[\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{x}_j)\Big]\Big) - L'(1)\mathrm{H}\left(\mathsf{f}_0 \circledast \left[\mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{f}_0)\right]\right)$$
$$- \mathrm{H}\Big(L^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{x}_j)\Big)\Big) + \mathrm{H}\left(L^{\circledast}(\mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{f}_0))\right).$$

It is easy to see that $F(\underline{\mathsf{f}_0}) = 0$, where $\underline{\mathsf{f}_0} = [\mathsf{f}_0, \cdots, \mathsf{f}_0]$. For fixed $\underline{\mathsf{x}} \succ \underline{\mathsf{f}_0}$, define $\phi\colon [0,1] \to \mathbb{R}$ as $\phi(t) = F(\underline{\mathsf{f}_0} + t(\underline{\mathsf{x}} - \underline{\mathsf{f}_0}))$. Then, $\phi(0) = F(\underline{\mathsf{f}_0})$, $\phi(1) = F(\underline{\mathsf{x}})$ and for $t \in [0,1]$,

$$\begin{aligned}
\phi'(t) &= \mathrm{d}_{\underline{\mathsf{x}}_1} F(\underline{\mathsf{x}}_1)[\underline{\mathsf{x}} - \underline{\mathsf{f}_0}]\,\Big|_{\underline{\mathsf{x}}_1 = \underline{\mathsf{f}_0} + t(\underline{\mathsf{x}} - \underline{\mathsf{f}_0})} \\
&= \frac{L'(1)}{w}\sum_{i=0}^{w-1} \mathrm{H}\Bigg(\Big[\mathsf{f}_0 - \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \boxtimes \rho^{\boxtimes}(t\mathsf{x}_j + (1-t)\mathsf{f}_0)\Big)\Big] \\
&\qquad\qquad \circledast \Big[\mathsf{c} \boxtimes \rho'^{\boxtimes}(t\mathsf{x}_i + (1-t)\mathsf{f}_0) \boxtimes (\mathsf{x}_i - \mathsf{f}_0)\Big]\Bigg) \\
&= \frac{L'(1)}{w}\sum_{i=0}^{w-1} \mathrm{H}\Bigg(\Big[\lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \boxtimes \rho^{\boxtimes}(t\mathsf{x}_j + (1-t)\mathsf{f}_0)\Big) - \mathsf{f}_0\Big] \\
&\qquad\qquad \boxtimes \Big[\mathsf{c} \boxtimes \rho'^{\boxtimes}(t\mathsf{x}_i + (1-t)\mathsf{f}_0) \boxtimes (\mathsf{x}_i - \mathsf{f}_0)\Big]\Bigg).
\end{aligned}$$

Also, since $\underline{\mathsf{x}} \succ \underline{\mathsf{f}_0}$, $\mathsf{x}_i \succeq \mathsf{f}_0$ and $t\mathsf{x}_j + (1-t)\mathsf{f}_0 \succeq \mathsf{f}_0$. Thus,

$$\lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \boxtimes \rho^{\boxtimes}(t\mathsf{x}_j + (1-t)\mathsf{f}_0)\Big) \succeq \lambda^{\circledast}\Big(\frac{1}{w}\sum_{j=0}^{w-1} \mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{f}_0)\Big) = \lambda^{\circledast}(\mathsf{c} \boxtimes \rho^{\boxtimes}(\mathsf{f}_0)) = \mathsf{f}_0,$$

since f_0 is a fixed point. By Proposition 8(iii), $\phi'(t) \leq 0$. Thus, $\phi(1) \leq \phi(0)$, which implies $F(\underline{\mathsf{x}}) \leq F(\underline{\mathsf{f}_0}) = 0$ for any $\underline{\mathsf{x}} \succ \underline{\mathsf{f}_0}$. Consequently,

$$U_{\mathrm{c}}(\mathsf{S}(\underline{\mathsf{x}}); \mathsf{c}) - U_{\mathrm{c}}(\underline{\mathsf{x}}; \mathsf{c}) \leq U_{\mathrm{s}}(\mathsf{f}_0; \mathsf{c}) - U_{\mathrm{s}}(\mathsf{x}_{i_0}; \mathsf{c}).$$

II.H.6 Negativity of Potential Functional Beyond Potential Threshold

In this section, we discuss negativity of the potential functional (Lemma 30) beyond the potential threshold when $\mathtt{h}^* = \mathtt{h}^{\text{stab}}$.

Suppose $\mathtt{h}^* = \mathtt{h}^{\text{stab}}$. Consider any $\mathtt{h} > \mathtt{h}^{\text{stab}}$ and observe that $\lambda'(0)\rho'(1)\mathfrak{B}(\mathtt{c}(\mathtt{h})) > 1$. For some $\mathtt{x} \in \mathcal{X}$, define $\phi\colon [0,1] \to \mathbb{R}$, $\phi(t) = U_{\text{s}}(\Delta_\infty + t(\mathtt{x} - \Delta_\infty); \mathtt{c}(\mathtt{h}))$. According to Proposition 16, note that ϕ is a polynomial in t, and $\phi(0) = 0$. By Lemma 23, since Δ_∞ is a fixed point of single system DE, $\phi'(0) = 0$. Moreover,

$$\begin{aligned}
\phi''(0) &= L'(1)\mathrm{H}\Big(\left[\mathtt{y} - \mathtt{c}(\mathtt{h}) \circledast \lambda'^{\circledast}\left(\rho^{\boxtimes}(\Delta_\infty)\right) \circledast \left[\rho'^{\boxtimes}(\Delta_\infty) \boxtimes \mathtt{y}\right]\right] \\
&\qquad\qquad \circledast \left[\rho'^{\boxtimes}(\Delta_\infty) \boxtimes \mathtt{y}\right] \Big), \quad \text{where } \mathtt{y} = \mathtt{x} - \Delta_\infty. \\
&= L'(1)\rho'(1)\mathrm{H}\left([\mathtt{y} - \lambda'(0)\rho'(1)\mathtt{c}(\mathtt{h}) \circledast \mathtt{y}] \circledast \mathtt{y}\right) \\
&= L'(1)\rho'(1)\mathrm{H}\left(\mathtt{x} \circledast \mathtt{x} - \lambda'(0)\rho'(1)\mathtt{c}(\mathtt{h}) \circledast \mathtt{x} \circledast \mathtt{x}\right).
\end{aligned}$$

For a family of BEC or binary input AWGN channels, we can choose $\mathtt{x} \in \mathcal{X}$ such that $\mathtt{x}^{\circledast 2} = \mathtt{c}(\mathtt{h})^{\circledast n}$ for any $n \in \mathbb{N}$. For such a choice of $\mathtt{x}$,

$$\phi''(0) = L'(1)\rho'(1)\mathrm{H}\left(\mathtt{c}(\mathtt{h})^{\circledast n} - \lambda'(0)\rho'(1)\mathtt{c}(\mathtt{h})^{\circledast n+1}\right) = \frac{L'(1)\rho'(1)}{(\lambda'(0)\rho'(1))^n}(f(n) - f(n+1)),$$

where $f(n) = (\lambda'(0)\rho'(1))^n \,\mathrm{H}\left(\mathtt{c}(\mathtt{h})^{\circledast n}\right)$. Since $\lambda'(0)\rho'(1)\mathfrak{B}(\mathtt{c}(\mathtt{h})) > 1$, by Proposition 12,

$$\lim_{n\to\infty} \frac{1}{n} \log f(n) = \lambda'(0)\rho'(1)\mathfrak{B}(\mathtt{c}(\mathtt{h})) > 1.$$

As such, $\lim_{n\to\infty} f(n) = \infty$, and thus there exists $m \in \mathbb{N}$ such that $f(m) < f(m+1)$. Thus, for a suitable choice of $\mathtt{x}$ such that $\mathtt{x}^{\circledast 2} = \mathtt{c}(\mathtt{h})^{\circledast m}$, we have $\phi''(0) < 0$. Since ϕ is a polynomial with $\phi(0) = \phi'(0) = 0$, there exists a $t \in (0,1]$ such that $\phi(t) = U_{\text{s}}(\Delta_\infty + t(\mathtt{x} - \Delta_\infty); \mathtt{c}(\mathtt{h})) < 0$. Thus, we have produced a suitable $\mathtt{x}$ for which $U_{\text{s}}(\mathtt{x}; \mathtt{c}(\mathtt{h})) < 0$. This completes the discussion for BEC and binary input AWGN channels.

For general BMS channels, we can show the same result under the condition

$$\lim_{n\to\infty} \frac{\mathrm{H}\left(\mathtt{x}^{\circledast n+1}\right)}{\mathrm{H}\left(\mathtt{x}^{\circledast n}\right)} = \mathfrak{B}(\mathtt{x}).$$

For this to hold, by Proposition 12, it suffices to show that the limit

$$\lim_{n\to\infty} \mathrm{H}\left(\mathsf{x}^{\circledast n+1}\right)/\mathrm{H}\left(\mathsf{x}^{\circledast n}\right)$$

exists. One way to guarantee the existence of such a limit is to show that the sequence of numbers $\{\mathrm{H}\left(\mathsf{x}^{\circledast n}\right)\}$ is log-convex,

$$\mathrm{H}\left(\mathsf{x}^{\circledast n+1}\right)\mathrm{H}\left(\mathsf{x}^{\circledast n-1}\right) \geq \mathrm{H}\left(\mathsf{x}^{\circledast n}\right)^2,$$

which itself follows by showing that the sequence $\{\mathrm{H}\left(\mathsf{x}^{\circledast n}\right)\}$ is completely monotonic [52, Proposition 4.7, Appendix A]. That is, the k-th differences of the sequence $\{\mathrm{H}\left(\mathsf{x}^{\circledast n}\right)\}$,

$$\mathrm{H}\left(\mathsf{x}^{\circledast n} \circledast (\mathsf{x}-\Delta_0)^{\circledast k}\right) = (-1)^k \mathrm{H}\left(\mathsf{x}^{\circledast n} \circledast (\Delta_0-\mathsf{x})^{\circledast k}\right),$$

have the sign $(-1)^k$. That first and second differences of this sequence have the sign -1 and $+1$, respectively, follows from Proposition 8. However, it remains to show

$$\mathrm{H}\left(\mathsf{x}^{\circledast n} \circledast (\Delta_0-\mathsf{x})^{\circledast k}\right) > 0, \quad \text{for } k > 2.$$

II.H.7 Connecting the Potential Functional and the RS Free Entropy

The purpose of this section is to provide pedagogical insight into the potential functional. As such, the following discussion is independent from the results of this chapter and the uninterested reader may skip this subsection.

The potential functional in Definition 20 can be viewed as a Lyapunov function. For the problem at hand, the negative of the replica-symmetric (RS) free entropy associated with the code ensemble is both a "natural" and an "optimal" Lyapunov function. It is optimal in the sense that it allows one to prove threshold saturation up to the MAP threshold (as $w \to \infty$), and it is natural because of its connection to RS formulas of statistical physics. Below, we first describe the RS free entropy for a general statistical mechanical system and then show how the corresponding expression for an LDPC ensemble reduces to the negative of the potential functional in Definition 20. We then briefly describe how the calculations change for LDGM ensembles. The choice of the negative sign for the potential is a convention for

consistency with [11, 12, 35].[4]

II.H.7.1 RS Free Entropy of General Graphical Models

Consider a graphical model on a bipartite graph $G = (V, C, E)$ with variable-node set V, a factor-node set C, and a set E of edges connecting variable- and factor-nodes. Let $\mathcal{A}$ be a discrete alphabet (for example $\mathcal{A} = \{0, 1\}$). Then, $\mathcal{A}^{|V|}$ is the set of all possible assignments to the variable-nodes. For $i \in V$, we denote the neighborhood of i ∂i as the set of all factor-nodes a such that $(i, a) \in E$; for $a \in C$, a similar definition is given for ∂a. For $\underline{x} \in \mathcal{A}^{|V|}$ and a subset $U \subset V$, we write $(x_i)_{i \in U}$ for the collection of elements in $\{x_i | i \in U\}$.

Each variable-node $i \in V$ has an associated weight function $g_i \colon \mathcal{A} \to [0, \infty)$, and each factor-node $a \in C$ has an associated function $f_a \colon \mathcal{A}^{|\partial a|} \to [0, \infty)$, which is a mapping from assignments of variable-nodes in ∂a, i.e. a function acting on unordered sets. One is generally interested in the marginals of the probability measure

$$P(\underline{x}) = \frac{1}{Z} \prod_{a \in C} f_a((x_i)_{i \in \partial a}) \prod_{i \in V} g_i(x_i),$$

where the normalizing factor

$$Z = \sum_{\underline{x} \in \mathcal{A}^{|V|}} \prod_{a \in C} f_a((x_i)_{i \in \partial a}) \prod_{i \in V} g_i(x_i)$$

is called the partition function. The free entropy is defined as

$$\frac{1}{|V|} \log Z.$$

The quantity $\log Z$ is closely related to the conditional entropy of the input in a communication channel given the output, and thus it naturally appears in a MAP decoding problem. See [53, Section 15.4] for more details.

It is well known that when G is a *tree*, a recursive evaluation of the sums allows one to solve for the marginals and the partition function exactly using the message

[4] This convention is also consistent with physics concepts: because parity-checks of LDPC codes are hard constraints, the RS free entropy is the negative of the RS free energy, thus the potential functional is the RS free energy. Moreover, in physics, entropies are maximized and energies, potentials are minimized.

passing formulas:

$$\mu_{i\to a}(x_i) = \frac{g_i(x_i)\prod_{b\in\partial i\setminus a}\hat{\mu}_{b\to i}(x_i)}{\sum_{x_i\in\mathcal{A}} g_i(x_i)\prod_{b\in\partial i\setminus a}\hat{\mu}_{b\to i}(x_i)}$$
$$\hat{\mu}_{a\to i}(x_i) = \frac{\sum_{(x_j)_{j\in\partial a\setminus i}} f_a((x_j)_{j\in\partial a})\prod_{j\in\partial a\setminus i}\mu_{j\to a}(x_j)}{\sum_{(x_j)_{j\in\partial a}} f_a((x_j)_{j\in\partial a})\prod_{j\in\partial a\setminus i}\mu_{j\to a}(x_j)}.$$

On a tree, these formulas are solved by initializing the messages emanating from leaf nodes and then recursively computing all the other messages. When a leaf node is the factor-node a, the outgoing message is $\hat{\mu}_{a\to i}(x_i)\propto f_a(x_i)$. Note that the factor-node degree is one here. When it is a variable-node i, the outgoing message is $\mu_{i\to a}(x_i)\propto g_i(x_i)$. The marginal distribution μ_i at variable-node $i\in V$ is then given by

$$\mu_i(x_i) = \frac{g_i(x_i)\prod_{a\in\partial i}\hat{\mu}_{a\to i}(x_i)}{\sum_{x_i\in\mathcal{A}} g_i(x_i)\prod_{a\in\partial i}\hat{\mu}_{a\to i}(x_i)}.$$

The free entropy on a tree is given by the Bethe formula

$$\frac{1}{|V|}\log Z = \frac{1}{|V|}\Big(\sum_{i\in V}\varphi_i + \sum_{a\in C}\phi_a - \sum_{(i,a)\in E}\psi_{i,a}\Big), \tag{II.17}$$

where

$$\varphi_i \triangleq \log\Big(\sum_{x_i\in\mathcal{A}} g_i(x_i)\prod_{b\in\partial i}\hat{\mu}_{b\to i}(x_i)\Big),$$
$$\phi_a \triangleq \log\Big(\sum_{(x_i)_{i\in\partial a}} f_a((x_i)_{i\in\partial a})\prod_{j\in\partial a}\hat{\mu}_{j\to a}(x_j)\Big),$$
$$\psi_{i,a} \triangleq \log\Big(\sum_{x_i\in\mathcal{A}}\mu_{i\to a}(x_i)\hat{\mu}_{a\to i}(x_i)\Big).$$

When G is not a tree, it is usually difficult to calculate the free entropy exactly. In this case, (II.17) can be seen as the pseudo-dual of the Bethe free entropy [54]. It also provides a first, a priori uncontrolled, approximation for the free entropy.

We now concentrate on *random graphical models* where G is an instance of a random bipartite graph. We assume that the functions f_a and g_i are realizations of possibly random functions f and g. For example, the weight function $g_i(x_i;Y_i)$ could

be an implicit function of random observation Y_i. An application to LDPC ensembles below will make this framework clear. Also, we denote by $\mathbb{E}[\cdot]$, the expectation with respect to all random objects.

The RS free entropy functional is an average of the Bethe formula (II.17) applied to the graph ensemble. Fix a trial probability measure m over the simplex

$$\left\{ (\alpha_1, \cdots, \alpha_{|\mathcal{A}|}) \in [0,1]^{|\mathcal{A}|} \mid \sum_i \alpha_i = 1 \right\}.$$

Let $\mu = (\mu(x))_{x \in \mathcal{A}}$ be a random variable distributed according to m, where the random variables $\mu(x)$, for $x \in \mathcal{A}$, are its components. Draw an integer r_e from the *edge-perspective* factor-node degree distribution. Let μ_i, for $i = 1, \cdots, r_e - 1$, be iid random variables distributed according to m. In the following, we define a new random variable $\hat{\mu}$, over the simplex given above, by its components:

$$\hat{\mu}(x) \triangleq \frac{\displaystyle\sum_{(x_1,\ldots,x_{r_e-1}) \in \mathcal{A}^{r_e-1}} f_a(x, x_1, \cdots, x_{r_e-1}) \prod_{i=1}^{r_e-1} \mu_i(x_i)}{\displaystyle\sum_{(x_0,x_1,\cdots,x_{r_e-1}) \in \mathcal{A}^{r_e}} f_a(x_0, x_1, \cdots, x_{r_e-1}) \prod_{i=1}^{r_e-1} \mu_i(x_i)}.$$

Draw integers r, ℓ from the *node-perspective* factor- and variable-node degree distributions, respectively. Let μ_i for $i = 1, \cdots, r$ and $\hat{\mu}_i$ for $i = 1, \cdots, \ell$ be independent copies of μ and $\hat{\mu}$, respectively.

Define the RS free entropy functional, a function of the trial distribution m, as

$$\begin{aligned}
\Phi_{\mathrm{RS}}(\mathsf{m}) \triangleq\ & \mathbb{E}\Big[\log\Big(\sum_{x \in \mathcal{A}} g(x) \prod_{j=1}^{\ell} \hat{\mu}_j(x) \Big) \Big] \\
& + \frac{L'(1)}{R'(1)} \mathbb{E}\Big[\log\Big(\sum_{(x_1,\ldots,x_r) \in \mathcal{A}^r} f(x_1, \cdots, x_r) \prod_{i=1}^{r} \mu_i(x_i) \Big) \Big] \\
& - L'(1) \mathbb{E}\Big[\log\Big(\sum_{x \in \mathcal{A}} \mu(x) \hat{\mu}(x) \Big) \Big]. \qquad \text{(II.18)}
\end{aligned}$$

Each successive term is an average of the variable, factor and edge sums in the Bethe formula (II.17). We note that $\mathbb{E}[\ell] = L'(1)$ and $\mathbb{E}[r] = R'(1)$. The coefficient $L'(1)/R'(1)$ accounts for the average number of factor-nodes per variable-node in the second term, and $L'(1)$ accounts for the average number of edges per variable-node

in the third term.

The RS approximation for the free entropy of a random graphical model is given by the minimum of this functional over an appropriate class of trial measures m. This approximation, or it's more sophisticated versions, may or may not be exact. Exactness of the RS formulas, if true, is usually difficult to prove and is the subject of various conjectures.

Finally, we point out that such formulas for sparse graph models were first derived in the framework of the replica method [55]. Apart from the conceptual problems related to the replica method, the derivations are also quite algebraically involved for the case of sparse graphs. The approach presented here via the Bethe formalism is better suited to sparse graphs and is of a more probabilistic nature.

II.H.7.2 Application to LDPC ensembles

We now specialize the RS free entropy functional to the LDPC(λ, ρ) ensemble. Here, the alphabet is binary, $\mathcal{A} \in \{0, 1\}$. The quantity $P(\underline{x})$ is the posterior probability of the input vector given the output vector. The parity check constraint functions are $f_a((x_i)_{i\in\partial a}) = \mathbf{1}(\oplus_{i\in\partial a} x_i = 0)$, and the weight function at a variable-node is the prior from channel observations, $g_i(x_i) = \Pr(Y_i|x_i)/\Pr(Y_i|0) = e^{-l_i x_i}$, where l_i is the LLR of the memoryless channel output assuming that 0 was transmitted.[5]

Remark: It is instructive to note that it is possible to choose different functions g_i without changing $P(\underline{x})$, e.g. $g_i(x_i) = e^{l_i(1-2x_i)/2}$ is chosen in [36, 42]. Depending on the choice of g_i, the Bethe free entropy may be different. However, the estimate of the conditional entropy can be adjusted accordingly and remains independent of the choice of the functions g_i.

Since the alphabet is binary, we can parameterize the vectors $(\mu(0), \mu(1))$ and $(\hat{\mu}(0), \hat{\mu}(1))$ by real valued random variables ν and $\hat{\nu}$ as follows:

$$\nu = \log \frac{\mu(0)}{\mu(1)}, \qquad\qquad \hat{\nu} = \log \frac{\hat{\mu}(0)}{\hat{\mu}(1)}.$$

Equivalently,

$$\mu(x) = \frac{1 + (-1)^x \tanh \frac{\nu}{2}}{2}, \qquad\qquad \hat{\mu}(x) = \frac{1 + (-1)^x \tanh \frac{\hat{\nu}}{2}}{2}.$$

[5]The random variable l_i is distributed according to the BMS channel c.

The random variable ν is distributed according to a trial measure n. By taking $r_e - 1$ independent copies $\nu_1, \cdots, \nu_{r_e-1}$ of ν, it is easy to show that $\hat{\nu}$ has the same distribution as

$$\hat{\nu} \sim 2 \tanh^{-1} \left(\prod_{i=1}^{r_e-1} \tanh \tfrac{\nu_i}{2} \right). \tag{II.19}$$

Also, take r independent copies $\nu_1, \cdots, \nu_r$ of ν, and ℓ independent copies $\hat{\nu}_1, \cdots, \hat{\nu}_\ell$ of $\hat{\nu}$. Straightforward algebra shows that the RS free entropy functional in (II.18) is given by

$$\begin{aligned} \Phi_{\mathrm{RS,LDPC}}(\mathsf{n}) = \mathbb{E}\Big[\log \Big(\prod_{j=1}^{\ell} \frac{1}{2} \Big[1 + \tanh \tfrac{\hat{\nu}_j}{2} \Big] + e^{-l} \prod_{j=1}^{\ell} \frac{1}{2} \Big[1 - \tanh \tfrac{\hat{\nu}_j}{2} \Big] \Big) \Big] \\ + \frac{L'(1)}{R'(1)} \mathbb{E}\Big[\log \Big(\frac{1}{2} \Big[1 + \prod_{i=1}^{r} \tanh \tfrac{\nu_i}{2} \Big] \Big) \Big] \\ - L'(1) \mathbb{E}\Big[\log \Big(\frac{1}{2} \Big[1 + \tanh \tfrac{\nu}{2} \tanh \tfrac{\hat{\nu}}{2} \Big] \Big) \Big], \end{aligned} \tag{II.20}$$

where the random variable l is distributed according to the BMS channel c. We note that the above expectation $\mathbb{E}[\cdot]$ includes the average over the LDPC(λ, ρ) ensemble via the integers ℓ and r drawn according to the variable- and check-node degree distributions, respectively.

We will now relate (II.20) to the potential functional in Definition 20. First note that the definitions of the operators $\circledast$ and $\boxast$ in Section II.B imply for any $k \geq 1$ and symmetric measures x_i, $i = 1, \ldots, k$,

$$\mathrm{H}\left(\circledast_{i=1}^{k} \mathsf{x}_i \right) = \int \log_2 (1 + e^{-\sum_{i=1}^{k} \alpha_i}) \prod_{i=1}^{k} \mathsf{x}_i(d\alpha_i), \tag{II.21}$$

$$\mathrm{H}\left(\boxast_{i=1}^{k} \mathsf{x}_i \right) = - \int \log_2 \Big(\frac{1}{2} \Big[1 + \prod_{i=1}^{k} \tanh \tfrac{\alpha_i}{2} \Big] \Big) \prod_{i=1}^{k} \mathsf{x}_i(d\alpha_i). \tag{II.22}$$

First consider the second term in (II.20). Using (II.22), since ν_i is distributed ac-

cording to n,

$$\frac{L'(1)}{R'(1)}\mathbb{E}\Big[\log\Big(\frac{1}{2}\Big[1+\prod_{i=1}^{r}\tanh\tfrac{\nu_i}{2}\Big]\Big)\Big]=-(\log 2)\frac{L'(1)}{R'(1)}\mathrm{H}\left(R^{\boxtimes}(\mathsf{n})\right). \tag{II.23}$$

For the third term in (II.20), since $\hat{\nu}$ is distributed according to (II.19), using (II.22),

$$\begin{aligned}
&L'(1)\mathbb{E}\Big[\log\Big(\frac{1}{2}\Big[1+\tanh\tfrac{\nu}{2}\tanh\tfrac{\hat{\nu}}{2}\Big]\Big)\Big]\\
&=L'(1)\mathbb{E}\Big[\log\Big(\frac{1}{2}\Big[1+\tanh\tfrac{\nu}{2}\prod_{i=1}^{r_e-1}\tanh\tfrac{\nu_i}{2}\Big]\Big)\Big]\\
&=-(\log 2)L'(1)\mathrm{H}\left(\mathsf{n}\boxtimes\rho^{\boxtimes}(\mathsf{n})\right).
\end{aligned} \tag{II.24}$$

For the first term in (II.20), we have

$$\begin{aligned}
&\mathbb{E}\Big[\log\Big(\prod_{j=1}^{\ell}\frac{1}{2}\Big[1+\tanh\tfrac{\hat{\nu}_j}{2}\Big]+e^{-l}\prod_{j=1}^{\ell}\frac{1}{2}\Big[1-\tanh\tfrac{\hat{\nu}_j}{2}\Big]\Big)\Big]\\
&=\mathbb{E}\Big[\sum_{j=1}^{\ell}\log\Big(\tfrac{1}{2}\Big[1+\tanh\tfrac{\hat{\nu}_j}{2}\Big]\Big)\Big]+\mathbb{E}\Big[\log(1+e^{-l-\sum_{j=1}^{\ell}\hat{\nu}_j})\Big]\\
&=L'(1)\mathbb{E}\Big[\log\Big(\tfrac{1}{2}\Big[1+\tanh\tfrac{\hat{\nu}}{2}\Big]\Big)\Big]+\mathbb{E}\Big[\log(1+e^{-l-\sum_{j=1}^{\ell}\hat{\nu}_i})\Big]\\
&=-(\log 2)L'(1)\mathrm{H}\left(\rho^{\boxtimes}(\mathsf{n})\right)+(\log 2)\mathrm{H}\left(\mathsf{c}\circledast L^{\circledast}(\rho^{\boxtimes}(\mathsf{n}))\right),
\end{aligned} \tag{II.25}$$

where we used (II.19), (II.22) and (II.21) to get the last equality.

Collecting (II.25), (II.23), (II.24), we find that

$$\Phi_{\mathrm{RS,LDPC}}(\mathsf{n})=-(\log 2)U_{\mathrm{s}}(\mathsf{n};\mathsf{c}),$$

which shows that the potential functional is the negative of the RS free entropy functional.

For completeness, we point out that the conditional entropy $\mathrm{H}\left(X^n|Y^n\right)$ of the input X^n conditional on the output Y^n is equal to the free entropy averaged over the noise realizations $\mathbb{E}[\mathrm{H}\left(X^n|Y^n\right)]=\mathbb{E}[\log_2 Z]$. For a detailed discussion of this relation, see [36, 39, 42]. Again, we note that due to different normalizations of the free entropy, additional nuisance terms may appear in these references. As stated in

Lemma 32, it is shown in these references that

$$\mathbb{E}[\mathrm{H}\left(X^n|Y^n\right)] \geq -\inf_{\mathsf{x}\in\mathcal{X}} U_{\mathrm{s}}(\mathsf{x};\mathsf{c}(\mathtt{h})).$$

It is conjectured that this is in fact an equality, and recently the equality has been proven for a class of regular codes and smooth channel families [43]. This is a case where the replica formula allows an exact calculation of the average free entropy.

II.H.7.3 Application to LDGM ensembles

We now briefly describe the calculations involved in obtaining the potential functional for LDGM ensembles in Definition 49. Observing the Tanner graph representation of an LDGM code in Fig. II.4, each generator-node a is connected to a code-bit x_a, and to each code-bit x_a there is an associated observation l_a, which is the LLR of the channel output. The parity-check constraint function at the generator-node a is given by

$$f_a((u_i)_{i\in\partial a}) = e^{-l_a x_a}\mathbf{1}(\oplus_{i\in\partial a} u_i \oplus x_a = 0).$$

In the set ∂a above, we do not include the neighbor x_a. The weight function at an information-node is given by $g_i(u_i) = 1$.

With the above functions, the RS free entropy in (II.18) for LDGM ensembles is given by

$$\begin{aligned}\Phi_{\mathrm{RS,LDGM}}(\mathsf{n}) = \mathbb{E}\Big[\log\Big(\prod_{j=1}^{\ell}\frac{1}{2}\Big[1+\tanh\tfrac{\hat{\nu}_j}{2}\Big] + \prod_{j=1}^{\ell}\frac{1}{2}\Big[1-\tanh\tfrac{\hat{\nu}_j}{2}\Big]\Big)\Big] \\ + \frac{L'(1)}{R'(1)}\mathbb{E}\left[\log\left(\frac{1+\prod_{i=1}^{r}\tanh\frac{\nu_i}{2} + e^{-l}\Big[1-\prod_{i=1}^{r}\tanh\frac{\nu_i}{2}\Big]}{2}\right)\right] \\ - L'(1)\mathbb{E}\Big[\log\Big(\frac{1}{2}\Big[1+\tanh\tfrac{\nu}{2}\tanh\tfrac{\hat{\nu}}{2}\Big]\Big)\Big], \qquad \text{(II.26)}\end{aligned}$$

where the random variable l is distributed according to c, and $\hat{\nu}$ has the same distribution as $\hat{\nu} \sim 2\tanh^{-1}\left(\tanh\frac{l}{2}\prod_{i=1}^{r_e-1}\tanh\frac{\nu_i}{2}\right)$. Proceeding as in the LDPC case, the three terms in (II.26) are, respectively,

$$-(\log 2)L'(1)\mathrm{H}\left(\mathsf{c}\circledast\rho^{\circledast}(\mathsf{n})\right) + (\log 2)\mathrm{H}\left(L^{\circledast}(\mathsf{c}\circledast\rho^{\circledast}(\mathsf{n}))\right),$$

$$(\log 2)\frac{L'(1)}{R'(1)}\mathrm{H}\left(\mathsf{c}\right) - (\log 2)\frac{L'(1)}{R'(1)}\mathrm{H}\left(\mathsf{c} \circledast R^{\circledast}(\mathsf{n})\right), \quad (\log 2)L'(1)\mathrm{H}\left(\mathsf{n} \circledast \mathsf{c} \circledast \rho^{\circledast}(\mathsf{n})\right),$$

which gives the relation $\Phi_{\mathrm{RS,LDGM}}(\mathsf{n}) = -(\log 2)U_{\mathrm{s}}(\mathsf{n};\mathsf{c})$.

CHAPTER III

SPATIALLY-COUPLED CODES FOR WYNER-ZIV, GELFAND-PINSKER, WRITE-ONCE MEMORY SYSTEMS*

III.A INTRODUCTION

In this chapter, we focus on three coding problems: 1) rate distortion with side-information (Wyner-Ziv formulation) 2) channel coding with side information (Gelfand-Pinsker formulation), and 3) coding for a write-once memory (WOM) system.

For lossy compression with a fixed distortion constraint and side information at the receiver, the Wyner-Ziv rate is the minimum achievable rate for source encoding [56]. For channel coding with an input constraint and side information at the transmitter, the Gelfand-Pinsker rate is the maximum achievable information rate [57]. Source and channel coding problems with side information arise naturally in network information theory. Solutions to these problems often require the use of "binning", where a set of codewords is partitioned into separate bins such that each bin is a good source code or channel code [58].

Another system related to the Gelfand-Pinsker problem is coding for a write-once memory (WOM) system. In a typical flash storage system, each cell carries an electric charge that indicates the stored bit; a higher charge denotes a 1 and a lower charge denotes a 0. While the charge of each cell can be raised easily, decreasing the charge requires resetting a large block of cells. The WOM systems model such storage cells. For a binary WOM, a bit with a value of 1 cannot be set to 0. Soon after its introduction by Rivest and Shamir [59], the capacity region of the noiseless WOM system was given by Heegard [60].

Sparse-graph codes with iterative decoding now offer many low-complexity solutions to channel coding problems. This good news does not extend automatically,

however, to problems that require the quantization of an arbitrary sequence to a nearby codeword (e.g., lossy compression, Wyner-Ziv, and Gelfand-Pinsker). These problems are not easily solved via iterative decoding because the iterations only converge for a vanishing fraction of the entire sequence space. In the past few years, a number of practical constructions for these problems have been considered [61–64]. In all these cases, there are non-negligible performance gaps relative to what is achievable for related single-user problems.

Since their introduction, there has been a plethora of WOM-code constructions for noiseless systems [65–70]. These constructions are based on projective geometry [66], coset coding [67, 70], graph coverings [68], and position modulation [69]. However, the first capacity-achieving scheme for the noiseless two-write WOM system, with polynomial encoding and decoding complexity, was introduced only recently [71]. Shortly after, polar codes were constructed with an encoding and decoding complexity of $O(n \log(n))$ that achieve the capacity region of the noiseless t-write WOM system [72]. Constructions based on LDGM codes are also considered in [73]. In that work, a sequence of optimized irregular LDGM ensembles is used to achieve capacity for the second write of the noiseless 2-write WOM system. The literature on WOM codes that handle errors is more limited. Error-correcting WOM codes were first constructed in [74, 75]. New constructions that are triple-error-correcting are presented in [76]. Polar codes were shown to correct a constant fraction (of blocklength) of errors [77]. Recently, polar codes were used to achieve the capacity of t-write WOM systems with write errors [78].

In this chapter, we seek to construct low-complexity capacity achieving codes for the three coding problems described above. For lossy compression, low-density generator matrix (LDGM) codes with modified versions of belief-propagation (BP) decoding can achieve practically interesting results but they appear to have a non-negligible gap to the optimal compression rate [79,80]. Better results can be obtained by using guided decimation techniques where an iterative algorithm (e.g., BP or survey propagation) is used to sequentially identify bits with a large bias and then fix them to match the bias [81, 82]. Building on these results, spatially-coupled LDGM codes with BP guided decimation (BPGD) were recently shown by Aref et al. to approach the optimal compression rate for a binary symmetric source [48, 83].

We extend the approach of Aref et al. to the Wyner-Ziv and Gelfand-Pinsker problems defined by binary symmetric sources and binary symmetric channels. For

these problems, Wainwright and Martinian showed that compound LDGM/LDPC codes can achieve the optimal rates when maximum-likelihood processing is used [13]. In this work, we show empirically that BPGD of spatially-coupled compound LDGM/ LDPC codes can approach the optimal Wyner-Ziv and Gelfand-Pinsker rates. A new complication with compound LDGM/LDPC codes is that BPGD encoding now has the possibility of failure. For iterative encoding with LDGM codes, encoding always succeeds but the resulting distortion depends on the code and encoder details. For compound LDGM/LDPC codes, the compressed bits are also required to satisfy some parity checks and the iterative encoder is not guaranteed to find a valid encoding regardless of the target distortion. This is similar to what happens when BPGD is used for constraint satisfaction problems [84]. In fact, without spatial-coupling the BPGD encoder failed in every experiment.

Our coding scheme for binary WOM system is also based on compound LDGM/ LDPC codes. By spatially coupling these codes, we are able to obtain codes that achieve the capacity under low-complexity message-passing algorithms for the 2-write WOM system. A key insight from [73] is that encoding the message for the second write can be reduced to the binary erasure quantization problem (BEQ). This reduction allows an efficient encoding algorithm with a linear computational complexity. We note that, while our encoder is also based on the reduction to the BEQ problem, our code construction differs from the one in [73] in a few important ways. In particular, it allows for low-complexity decoding and it can handle errors in the read process. Moreover, by changing the encoding algorithm, we can also handle the first write.

Compound LDGM/LDPC codes are capacity achieving under MAP decoding [13, 85]. Also, when these codes are spatially-coupled, they also achieve the capacity under iterative decoding [86, 87]. As a consequence, our constructions based on spatially-coupled compound codes are also capacity achieving. Moreover, spatial coupling also obviates the need to optimize degree distributions to operate close to capacity.

Finally, it is important to note that polar codes also allow the deterministic construction of capacity achieving codes for all these problems [72,78,88]. Unfortunately, the finite-length performance of these constructions is not good. As such, they require very large blocklengths to operate at rates close to the capacity limit. The precise trade-offs between spatially-coupled codes and polar codes is still not clear

and thus remains as an interesting open problem.

III.B SYSTEM MODEL

In this section, we describe the three problems of interest in this chapter: 1) rate distortion with side information (Wyner-Ziv formulation), 2) channel coding with side information (Gelfand-Pinsker formulation), and 3) coding for a write-once memory system. We seek to construct coding schemes with low computational complexity that achieve the rate regions of these problems.

Notational convention:

- Addition modulo 2 is denoted by $\oplus$.
- A Bernoulli random variable with parameter δ is denoted by $\mathsf{Ber}(\delta)$.
- The binary entropy function is denoted by $h(x) \triangleq -x\log_2 x-(1-x)\log_2(1-x)$.

III.B.1 Rate Distortion with Side Information

Consider a linear code $\mathcal{C} \subset \{0,1\}^n$ with rate $R = k/n$ and an iid sequence $X^n = (X_1, \ldots, X_n) \in \{0,1\}^n$, where each X_i is a $\mathsf{Ber}(\frac{1}{2})$ random variable. In the rate distortion problem, the objective is to optimally encode X^n to a codeword $\widehat{X}^n \in \mathcal{C}$ that minimizes the expected normalized Hamming distortion D between X^n and $\widehat{X}^n$, $D = \frac{1}{n}\sum_{i=1}^n \mathbb{E}|X_i - \widehat{X}_i|$. For this problem, Shannon's rate-distortion theory [58] shows that any rate $R > 1-h(D)$ is both achievable and necessary. Thus, roughly $n(1-h(D))$ bits are required to specify X^n up to normalized distortion D.

In the Wyner-Ziv formulation of the rate distortion problem, there is side information Z^n at the decoder about X^n. Suppose the side information Z^n takes the form $Z_i = X_i \oplus \mathsf{Ber}(p)$. Then, [56] shows that any rate

$$R > R_{WZ}(D,p) = l.c.e\{h(D*p) - h(D), (p,0)\}, \tag{III.1}$$

is achievable and necessary to describe X^n (up to normalized distortion D) along with side information Z^n, where $D*p = D(1-p)+p(1-D)$ and $l.c.e$ denotes the lower convex envelope.

III.B.2 Channel Coding with Side Information

Suppose again that $\mathcal{C} \subset \{0,1\}^n$ is a linear code with rate $R = k/n$ and that $W^n \in \{0,1\}^n$ is distributed iid according to $\mathsf{Ber}(\delta)$. In the channel coding problem,

the objective is to encode a message $s^k \in \{0,1\}^k$ into a codeword $X^n \in \mathcal{C}$ so that the decoder can reliably estimate the message s^k from the decoder output $X^n \oplus W^n$. When the channel $\{W_i\}$ is distributed iid according to $\mathsf{Ber}(\delta)$, classical result in information theory states that any rate $R < 1 - h(\delta)$ is both necessary and sufficient.

In the Gelfand-Pinsker formulation, the output at the receiver is given by $Y^n = X^n \oplus Z^n \oplus W^n$, where the side information $Z^n \in \{0,1\}^n$ is available a priori to the encoder but not the decoder. Additionally, the average weight of the codeword X^n is required to satisfy

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] \leq \delta, \qquad\qquad p < \delta \leq \tfrac{1}{2}.$$

Under these conditions, [57] shows that the rate constraint

$$R < R_{GP}(\delta, p) = h(\delta) - h(p) \tag{III.2}$$

is both necessary and sufficient.

III.B.3 Write-Once Memory System

Consider a storage system with n cells, wherein each cell can store one bit $\{0,1\}$ of information.

Definition 62: In a binary WOM system, a cell with a value of 1 cannot be changed to 0. This is called the WOM constraint.

In a single write, a k-bit message $s^k \in \{0,1\}^k$ is encoded into a n-bit sequence x^n in a binary WOM system. Such a write is associated with a rate of k/n. We consider read errors, where it is required to decipher the message s^k from a *noisy* version of x^n. Below, we discuss the noiseless and noisy systems separately.

III.B.3.1 Noiseless Write-Once Memory Systems

Consider a noiseless binary WOM system initialized with zeros, and a successive representation of t messages with rates $R_1, \ldots, R_t$. The capacity region of the t-write system is given by [60]

$$\Big\{(R_1,\ldots,R_t) \mid 0 \leq R_i \leq h(\delta_i)\prod_{j=1}^{i-1}(1-\delta_j), 1 \leq i \leq t-1,$$

$$0 \leq R_t \leq \prod_{j=1}^{t-1}(1-\delta_j), 0 \leq \delta_i \leq 1\Big\}, \tag{III.3}$$

where $h(x) = -x\log_2 x - (1-x)\log_2(1-x)$.

The capacity region in (III.3) has an intuitive description. Define the normalized weight of a binary sequence as the number of ones in the sequence divided by its length. Consider the first write and suppose that the normalized weight of the sequence in the first write is $0 \leq \delta_1 \leq 1$. Then, the rate of the first write must satisfy $R_1 < h(\delta_1)$. Now, a fraction $1-\delta_1$ of the cells that are zeros can be utilized for subsequent writes. For the second write, if a further fraction $0 \leq \delta_2 \leq 1$ of these cells are allowed to set to 1, then the rate of the second write satisfies $R_2 < (1-\delta_1)h(\delta_2)$. Generalizing this argument gives the upper bound in (III.3) on the achievable rates.

The primary focus here is on the second write of the 2-write WOM system. Suppose the normalized weight of the sequence after the first write is δ. Then, we construct a coding scheme for the second write of the 2-write WOM system that achieves any rate

$$R < 1 - \delta. \tag{III.4}$$

III.B.3.2 Write-Once Memory Systems with Read Errors

Now, consider the case where there are read errors. Suppose a message s^k is encoded into a sequence x^n in a WOM system. However, the message will be decoded from $y^n \in \{0,1\}^n$, a noisy version of x^n. This models the bit-flips caused by read errors in a storage system. We assume that the errors are caused by a BSC with bit-flip probability p. That is, $y_i = x_i \oplus \mathsf{Ber}(p)$, where $\oplus$ denotes the modulo 2 addition and $\mathsf{Ber}(p)$ denotes the Bernoulli random variable with parameter p.

It is important to note that this error model is different from the write error model. In that model, the error occurs during the write and the reads occur without error. In this case, the encoder can achieve optimal performance by biasing the input distribution. In the read-error model, the write occurs without errors but the each read experiences independent errors. If sufficient error correction is used, then a single read can be used to recover the correct codeword with high probability and the system will have perfect knowledge of the memory state during rewrites. To the best of the authors' knowledge, the capacity region of the WOM system with read

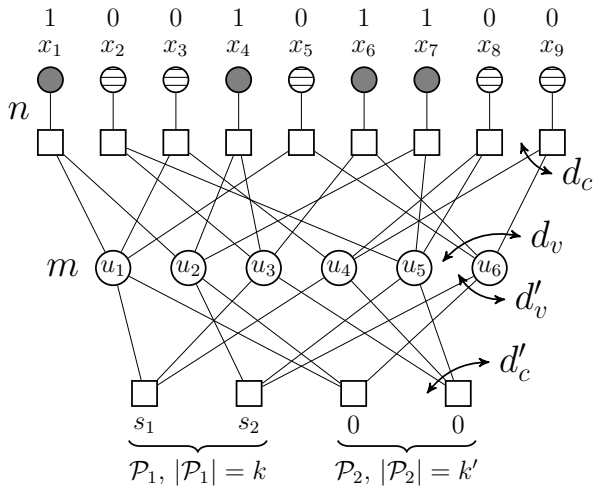

Figure III.1: A Tanner graph representation of a compound code. The top part represents the LDGM code, and the bottom part represents the LDPC code. The parities in $\mathcal{P}_1$ carry the message, and the parities in $\mathcal{P}_2$ provide the error correction.

errors is unknown.

For the second write of the two-write WOM system with read errors, we can construct codes that achieve any rate

$$R < 1 - \delta - h(p), \tag{III.5}$$

where δ is the normalized weight of the state sequence after first write.

III.C CODING SCHEME

III.C.1 Compound LDGM/LDPC Codes

We now describe coding schemes for the problems described in the previous section that achieve the rate regions in (III.1), (III.2), (III.4) and (III.5) under optimal encoding and decoding. The emphasis in the coding schemes described in Sections III.C.2, III.C.3, and III.C.4 is not on the computational complexity of the encoding and decoding. In Section III.C.5, we modify the coding scheme with spatially-coupled codes, and in Section III.D, we describe efficient message-passing algorithms for encoding and decoding that achieve the desired rate regions.

At the heart of these coding schemes is a compound LDGM/LDPC code [85], [4], [13]. We illustrate these codes through an example. The upper portion of the Tanner graph in Fig. III.1 represents a length-n LDGM code with information bit

length m, where the codeword is $x^n = (x_1, \ldots, x_n) \in \{0,1\}^n$ and the information bits are $(u_1, \ldots, u_m) \in \{0,1\}^m$. In the compound code, additionally, the information bits are required to satisfy certain parity constraints. These constraints are split into two groups $\mathcal{P}_1$ and $\mathcal{P}_2$ with $|\mathcal{P}_1| = k$ and $|\mathcal{P}_2| = k'$. The parities in $\mathcal{P}_1$ specify a fixed constraint given by $s^k = (s_1, \ldots, s_k) \in \{0,1\}^k$. For the example in Fig. III.1, this means $u_1 \oplus u_2 \oplus u_5 = s_1$ and $u_3 \oplus u_4 \oplus u_5 = s_2$. The parities in group $\mathcal{P}_2$ are the usual even parities and for the example, $u_1 \oplus u_3 \oplus u_6 = 0$, $u_2 \oplus u_4 \oplus u_6 = 0$. The bottom part therefore represents constraints akin to an LDPC code. As such, these codes are referred to as compound LDGM/LDPC codes. We distinguish between LDGM and LDPC check-nodes and we refer to the nodes representing x_i's as LDGM bit-nodes and u_i's as LDPC bit-nodes. The codebook is the set of all sequences x^n generated by the LDPC bit-nodes that satisfy the required constraints.

A few observations are in order. We note that the design rate of this code is

$$\frac{m-k-k'}{n} = \frac{d_c}{d_v}\left(1 - \frac{d_v'}{d_c'}\right)$$

The codebook specified by a compound code is linear if and only if the parity constraints in $\mathcal{P}_1$ satisfy $s^k = 0^k$. There is also a natural coset decomposition of these codes. Ignore for a moment the parity constraints specified by $\mathcal{P}_1$, and denote the resulting linear code by $\mathcal{C}$. Denote by $\mathcal{C}(s^k)$, the original codebook when the parities in $\mathcal{P}_1$ satisfy $s^k \in \{0,1\}^k$. It is easy to see that $\mathcal{C} = \bigcup_{s^k \in \{0,1\}^k} \mathcal{C}(s^k)$.

Now, we describe another codebook $\mathcal{C}'$ that is related to $\mathcal{C}$. The codebook $\mathcal{C}'$ is obtained from $\mathcal{C}$ by ignoring all the constraints in $\mathcal{P}_2$. In particular, $\mathcal{C}'$ is the linear code generated from the LDGM part of $\mathcal{C}$. In describing $\mathcal{C}$, the constraints in $\mathcal{P}_1$ are not active. When the constraints in $\mathcal{P}_1$ are active, the codebook generated by ignoring $\mathcal{P}_2$ is denoted by $\mathcal{C}'(s^k)$. Note that $\mathcal{C}'(s^k)$ is a compound code in itself (with an empty $\mathcal{P}_2$). We again have the coset decomposition of $\mathcal{C}'$ into $\mathcal{C}'(s^k)$.

The main attraction of the compound codes is that they are simultaneously good for rate distortion and channel coding. Here, we use the term *good* to mean the following.

Remark 63: We call a code "good for rate distortion" if the design rate of the code satisfies $R = 1 - h(D) + \varepsilon$ for a small ε, and when the code is used to optimally encode a $\mathsf{Ber}(\frac{1}{2})$ sequence, the average normalized Hamming distortion is at most D. Similarly, we call a code "good for channel coding" if $R = 1 - h(p) - \varepsilon$, and when the

code is used for the channel coding problem with the channel $\mathsf{Ber}(p)$, the message can be reliably estimated under optimal decoding with probability of error at most ε.

It is known that the codebooks $\mathcal{C}$, $\mathcal{C}(s^k)$, $\mathcal{C}'(s^k)$ are good for both rate distortion and channel coding for any $s^k \in \{0,1\}^k$ [13]. But $\mathcal{C}'$ is not a good channel code, since LDGM codes with fixed degrees exhibit non-negligible error floors. However, $\mathcal{C}'$ is good for rate distortion [82].

Below, we describe coding schemes for the three problems in Section III.B, using the codebooks $\mathcal{C}$, $\mathcal{C}'$, $\mathcal{C}(S)$, $\mathcal{C}'(S)$. That these schemes achieve the rate regions in (III.1) and (III.2) will be an immediate consequence of the fact that compound codes are simultaneously good for rate distortion and channel coding, and LDGM codes are good for rate distortion. We only make heuristic arguments in the following, but these can easily be made rigorous.

III.C.2 Coding Scheme for Wyner-Ziv

For small $\varepsilon > 0$, choose n, m, k such that

$$\frac{m}{n} = 1 - h(D) + \varepsilon/2, \qquad \frac{k}{n} = h(D * p) - h(D) + \varepsilon.$$

Consider a compound code with block length n, information bit length m, $|\mathcal{P}_1| = k$ and $|\mathcal{P}_2| = 0$ such that $\mathcal{C}'$ is a good rate distortion code with rate m/n and $\mathcal{C}'(s^k)$ is a good channel code with rate $1 - h(D * p) - \varepsilon/2$.

Suppose we want to encode $X^n \in \{0,1\}^n$ up to normalized distortion D, with side information Z^n in the form $Z_i = X_i \oplus \mathsf{Ber}(p)$. The sequence X^n can be encoded to $\widehat{X}^n \in \mathcal{C}'$ with distortion at most D, since $\mathcal{C}'$ is a good rate distortion code. The sequence $\widehat{X}^n$ belongs to a unique coset $\mathcal{C}'(\widehat{s^k})$ for some $\widehat{s^k} \in \{0,1\}^k$. The encoder transmits $\widehat{s^k}$ to the decoder, which requires a rate of $R = \frac{k}{n} = h(D * p) - h(D) + \varepsilon$.

The decoder, together with side information Z^n and $\widehat{s^k}$, tries to recover $\widehat{X}^n$. Note that

$$Z_i = X_i \oplus \mathsf{Ber}(p), \quad \text{and} \qquad \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}|X_i - \widehat{X}_i| \approx D.$$

This however implies $Z_i \approx \widehat{X}_i \oplus \mathsf{Ber}(D * p)$. Since $\mathcal{C}'(\widehat{s^k})$ is a good channel code with rate $1 - h(D * p) - \varepsilon/2$, the decoder can recover $\widehat{X}^n$ reliably. As such, any rate

$R > h(D * p) - h(D)$ is sufficient to describe X up to distortion D. Thus, the rate region in (III.1) is achievable with this scheme.

Remark 64: The coding scheme we described above is slightly different from [13]. The difference is in using $\mathcal{C}'$ instead of $\mathcal{C}$ when compressing X^n to $\widehat{X}^n$. The reason for our choice will be apparent later.

III.C.3 Coding Scheme for Gelfand-Pinsker

First, choose n, m, k, k' such that

$$\frac{m-k'}{n} = 1 - h(p) - \frac{\varepsilon}{2}, \qquad \frac{m-k-k'}{n} = 1 - h(\delta) + \frac{\varepsilon}{2}.$$

Consider a compound code with parameters n, m, k, k' such that $\mathcal{C}$ is a good channel code with rate $\frac{m-k'}{n}$ and $\mathcal{C}(s^k)$ is a good rate distortion code with rate $\frac{m-k-k'}{n}$.

We want to transmit a message $s^k \in \{0,1\}^k$, when the encoder has side information Z^n. First, at the encoder, the side information Z^n is compressed to $\widehat{Z}^n \in \mathcal{C}(s^k)$. Since $\mathcal{C}(s^k)$ is a good rate distortion code with rate $1 - h(\delta) + \varepsilon/2$, we have

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}|Z_i - \widehat{Z}_i| \approx \delta, \quad \text{or} \quad \widehat{Z}_i \approx Z_i \oplus \mathsf{Ber}(\delta).$$

The encoder transmits the vector $X^n = \widehat{Z}^n \oplus Z^n$. It is important to note that this choice of X^n has an average weight of δ. The output at the decoder is given by $Y^n = X^n \oplus Z^n \oplus W^n = \widehat{Z}^n \oplus W^n$, where the channel W^n is distributed iid according to $\mathsf{Ber}(p)$. Since $\mathcal{C}$ is a good channel code with rate $1 - h(p) - \varepsilon/2$, $\widehat{Z}^n$ can be reliably decoded at the receiver. Now, since $\widehat{Z}^n$ belongs to a unique coset $\mathcal{C}(s^k)$, the message s^k can be recovered reliably. Thus, any rate $R = k/n < h(\delta) - h(p)$ is achievable.

III.C.4 Coding Scheme for Write-Once Memory

We now describe our coding scheme for the second write of the 2-write WOM system based on compound codes. Let s^k be the k-bit message for the second write and n be the size of the WOM system. Consider Fig. III.1, where the gray LDGM bit-nodes represent the indices that are set to 1 and the horizontally patterned LDGM bit-nodes represent the zeroes after the first write. The parities s^k in group $\mathcal{P}_1$ carry the message.

Construction 65: The k-bit message s^k is encoded into a n-bit sequence x^n in the WOM system by finding a codeword x^n in the codebook $\mathcal{C}(s^k)$ with the constraint that the indices that are 1 after the first write remain the same in x^n.

For such a construction, the message s^k can then be retrieved from x^n (or a noisy version of x^n), which gives a rate of $R = k/n$. It remains to find codes $\mathcal{C}(s^k)$ and algorithms that allow Construction 65 and also achieve the rates in (III.4) and (III.5).

A crucial step in Construction 65 is finding a codeword in a code with specified values in certain bit positions. This is an instance of the *erasure quantization* problem. In the binary erasure quantization (BEQ) problem, it is required to quantize a source sequence over the alphabet $\{0, 1, *\}$ to some codeword in a given codebook over the binary field $\{0, 1\}$. The requirement is that 0s and 1s in the source sequence cannot be changed but the erasures $*$ can be set to either 0 or 1 in the codeword. To map the WOM problem to the BEQ problem, we can take the source sequence as the state of the WOM system after first write and set all zeroes to erasures.

A key observation in [89] is that the BEQ problem is closely related to the channel coding over the binary erasure channel (BEC). In particular, [89, Theorem 2] states that an erasure quantizer for a code can successfully quantize every source sequence with erasure pattern $e^n \in \{0, 1\}^n$ if and only if the channel decoder for the dual code can correct all received sequences with the erasure pattern $1^n \oplus e^n$.

Remark 66: It is well known that the compound codes achieve the capacity on erasure channels under bit-MAP decoding [85]. However, under bit-MAP decoding, linear codes achieve capacity if and only if their dual codes achieve capacity. Thus, the compound codes also achieve the capacity region of the BEQ problem.

Another important property of compound codes is that they are good channel codes under MAP decoding for the BSC [82]. Roughly speaking, this means that if a compound code (of large enough degrees and blocklength) of rate $1 - h(p) - \varepsilon$ is used for the channel coding problem over a BSC with parameter p, then the message can be reliably estimated under optimal decoding with probability of error at most ε.

Theorem 67: Compound codes achieve the rate regions in (III.4) and (III.5) under optimal encoding and decoding.

Proof. The following provides heuristic arguments that can be made rigorous using standard techniques. Moreover, we only discuss the achievability of (III.5) since (III.4) is a special of (III.5) with $p = 0$.

Choose a compound code $\mathcal{C}(s^k)$ with parameters $m = n$ and $(n - k')/n = 1 - h(p) - \varepsilon$. The code $\mathcal{C}$ is also a compound code with rate $(n - k)/n = 1 - h(p) - \varepsilon$. Since $\mathcal{C}$ is a good channel code for the BSC, the codeword x^n (and subsequently s^k) can be recovered from the BSC channel with parameter p.

Since the dual code $\mathcal{C}(s^k)^\perp$ (with rate $1 - (n - k - k')/n$) can correct up to a fraction $(n - k - k')/n$ of erasures, the code $\mathcal{C}(s^k)$ can quantize all source sequences with at least a fraction $1 - (n - k - k')/n$ of erasures. Thus, for a fraction of at most $(n - k - k')/n$ ones in the WOM system after first write, we can find a codeword in $\mathcal{C}(s^k)$ according to Construction 65. Thus, any $\delta < (n - k - k')/n = 1 - h(p) - R - \varepsilon$ is achievable, which is the rate region in (III.4). □

III.C.5 Spatially-Coupled Compound LDGM/LDPC Codes

We now describe coding schemes for our three problems based on spatially-coupled compound LDGM/LDPC codes. The encoding and decoding for these codes are based on practically implementable, polynomial time message-passing algorithms. In essence, we describe new codebooks $\mathcal{SC}$, $\mathcal{SC}(s^k)$, $\mathcal{SC}'$, $\mathcal{SC}'(s^k)$, analogous to $\mathcal{C}$, $\mathcal{C}(s^k)$, $\mathcal{C}'$, $\mathcal{C}'(s^k)$, that are good for rate distortion, channel coding, and binary erasure quantization, when encoding and decoding is done with message-passing algorithms [48,83,86,87]. The coding scheme for our problems is same as in the previous section except that we use spatially-coupled codes and use efficient message-passing algorithms for encoding and decoding. Below, we give the construction of spatially-coupled compound codes and discuss the message-passing algorithms in the next section.

For a comprehensive introduction to the spatially-coupled codes, see [10, 18, 19, 90]. Below, we give a brief description of the construction of the spatially-coupled compound LDGM/LDPC codes. We describe the construction with regular degrees for both bit- and check-nodes. As in Fig. III.1, let d_v and d'_v denote the LDPC bit-node degrees to the LDGM and LDPC check-nodes, respectively. Also, let d_c and d'_c denote the LDGM and LDPC check-node degrees to the LDPC bit-nodes. For simplicity, we assume that the parities in $\mathcal{P}_1$ have the same degree as the parities in $\mathcal{P}_2$.

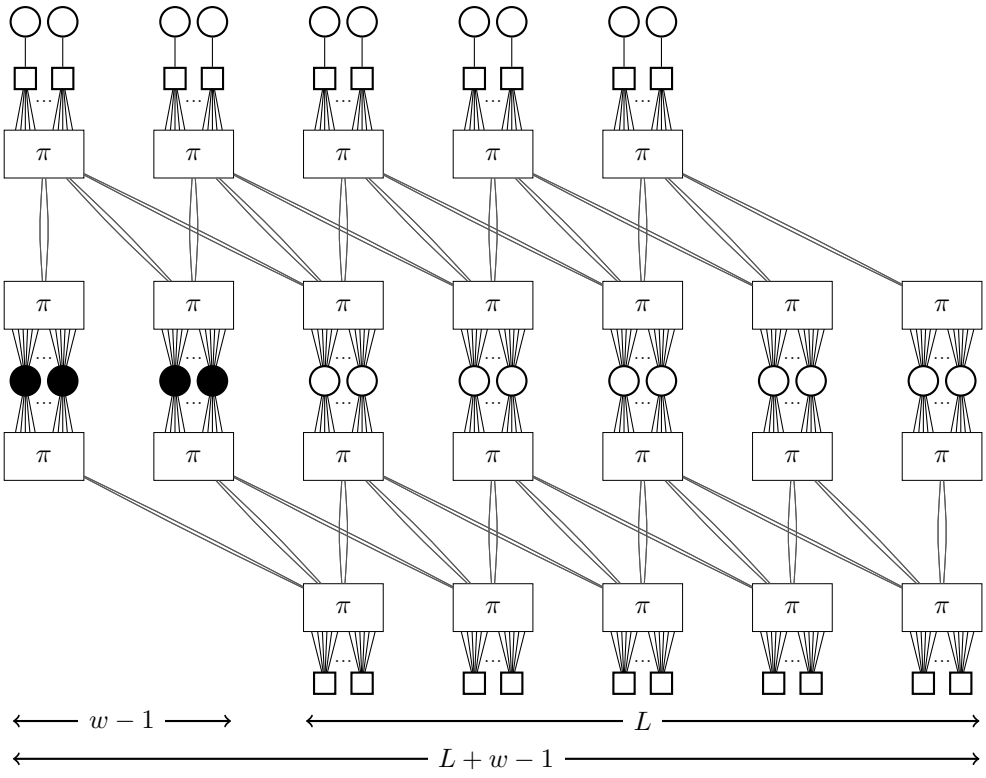

Figure III.2: Illustration of connections in a spatially-coupled compound LDGM/ LDPC code. The top part denotes the coupling in the LDGM part, and the bottom part denotes the coupling in the LDPC part. The LDPC bit-nodes in the first $w-1$ sections (black bit-nodes in the middle) are set to 0.

Let L be the *chain length* of the spatially-coupled system and w be the *coupling width* of the spatially-coupled code. Consider groups of LDPC bit-nodes at sections indexed by $\mathcal{N}_v = \{1, \cdots, L+w-1\}$, LDGM check-nodes at $\mathcal{N}_c = \{1, \cdots, L\}$, and LDPC check-nodes at $\mathcal{N}_c' = \{w, \cdots, L+w-1\}$. We describe the coupling structure in the LDGM part; similarly, this can be repeated for the LDPC part, both for the parities in $\mathcal{P}_1$ and $\mathcal{P}_2$. To begin with, let N denote the number of LDPC bit-nodes in a given section $i \in \mathcal{N}_v$, and pick N so that

- Nd_v/d_c, Nd_v'/d_c' are integers; this is to ensure that all LDGM and LDPC check-nodes have regular degrees,
- Nd_v/w, Nd_v'/w are integers; this ensures an appropriate partition of the edges in a given section for the coupling.

Place d_v edge sockets to each LDPC bit-node. Place Nd_v/d_c LDGM check-nodes

at each section in $\mathcal{N}_c$, each check-node with d_c edge sockets. For both LDPC bit-nodes and LDGM check-nodes, this ensures each a total of Nd_v edge sockets. Now, partition the Nd_v sockets at both LDPC bit-nodes and LDGM check-nodes into w groups using a uniform random permutation. Denote these partitions by $\mathcal{N}^v_{i,\ell}$, $\mathcal{N}^c_{j,\ell}$, where $1 \leq i \leq L+w-1$, $1 \leq j \leq L$, $1 \leq \ell \leq w$. The coupled LDGM component is constructed by connecting the sockets in $\mathcal{N}^c_{j,\ell}$ to $\mathcal{N}^v_{j+\ell-1,\ell}$. See Fig. III.2 for an illustration of these connections.

To construct the coupled LDPC component, one can start by placing d'_v edge sockets at each LDPC bit-node and repeating the above process. Moreover, the coupling is done separately for parities in $\mathcal{P}_1$ and $\mathcal{P}_2$, and ensure that each LDPC bit-node has connections to check-nodes in both $\mathcal{P}_1$ and $\mathcal{P}_2$. When decoding the message s^k from x^n, the channel decoder does not use the parities in $\mathcal{P}_1$. Thus, if there are LDPC bit-nodes with no connections to the parities in $\mathcal{P}_2$ (but has connections only to $\mathcal{P}_1$), this causes a small error floor in the decoding.

This construction leaves some edge sockets of the LDPC bit-nodes unconnected at the boundary. Similarly, the LDPC bit-nodes in the first $w-1$ sections are shortened to 0. The shortening and the boundary connections are necessary for the spatially-coupled "encoding/decoding wave" to get started and the threshold saturation phenomenon to take effect in spatial-coupling [10]. This termination results in a rate-loss that is a well-known side-effect of spatially-coupled constructions.

III.D MESSAGE-PASSING ALGORITHMS

Message-passing algorithms for channel coding are now a standard part of the coding theory literature. We refer the reader to [38] for their description. It has been recently shown [87], [86] that spatially-coupled compound LDGM/LDPC codes are good for channel coding under message-passing. As such, $\mathcal{SC}$, $\mathcal{SC}(s^k)$, $\mathcal{SC}'(s^k)$ are good for channel coding under message-passing.

In the following, we focus exclusively on the message-passing algorithm for rate distortion and binary erasure quantization. For rate distortion, we describe a variation of the so-called belief-propagation guided decimation (BPGD) algorithm [83], [48]. While there is no satisfactory theoretical analysis of the BPGD algorithm, numerical results shown in the next section confirm that for spatially-coupled compound LDGM/LDPC codes under the BPGD algorithm is good for rate distortion [48]. For the BEQ problem, we describe the so-called iterative quantization algorithm [89].

Algorithm 1 Belief-Propagation Guided Decimation

Require: Sequence $x^n \in \{0,1\}^n$ to encode, parameters $(T,\ \beta)$, graph $G(V,U,C)$.
Set $m_{i\to a}=\widehat{m}_{a\to i}=0$ for $i \in V \cup U$, $a \in C$ and $(i,a) \in G$.
Initialize set $V_{\rm dec}$ to equal LDPC bit-nodes in first $w-1$ sections.
while $V_{\rm dec} \neq V$ **do**
for $t = 1$ to T **do**
$m_{i\to a} = (-1)^{x_i} \tanh(\beta)$ for $i \in U$ and $a \in C$.
$m_{i\to a} = (-1)^{u_i} \cdot \infty$ for $i \in V_{\rm dec}$ and $a \in C$.
$m_{i\to a} = \sum_{b\in\partial i\setminus\{a\}} \widehat{m}_{b\to i}$ for $i \in V \setminus V_{\rm dec}$ and $a \in C$.
$\widehat{m}_{a\to i} = \tanh^{-1} \prod_{j\in\partial a\setminus\{i\}} \tanh m_{j\to a}$ for $i \in V \setminus V_{\rm dec}$ and $a \in C$.
end for
Evaluate $m_i = \sum_{a\in\partial i} \widehat{m}_{a\to i}$ for all $i \in V \setminus V_{\rm dec}$.
Set B to be max. of $|m_i|$ when i varies over left-most w sections of $V \setminus V_{\rm dec}$; denote the resulting bit-node by i^*.
if $B = 0$ **then**
Pick a bit-node i^* uniformly in left-most w sections of $V \setminus V_{\rm dec}$ and set u_{i^*} to be 0 or 1 uniformly randomly.
else
Set u_{i^*} to 0 or 1 with prob. $\frac{1+\tanh m_{i^*}}{2}$ or $\frac{1-\tanh m_{i^*}}{2}$.
end if
Set $V_{\rm dec} = V_{\rm dec} \cup \{i^*\}$.
end while
If $\{u_i\}$ fail to satisfy LDPC checks, then **repeat**.

That spatially-coupled compound LDGM/LDPC codes under iterative quantization algorithm achieve the capacity region of the BEQ problem is a consequence of the fact that these codes achieve the capacity region of the binary erasure channel. This equivalence is due to the duality between the binary erasure quantization and the channel coding under binary erasure channel.

III.D.1 Belief-Propagation Guided Decimation

Consider an instance of the spatially-coupled compound LDGM/LDPC described above. Denote its Tanner graph by $G(V,U,C)$, where V denotes the LDPC bit-nodes, U denotes the LDGM bit-nodes, C denotes the check-nodes (both LDGM and LDPC). Place a sequence $x^n \in \{0,1\}^n$ at the top of LDGM bit-nodes as in Fig. III.1. The message-passing rules here are same as in the channel coding setup by assuming that x_i have come through a BSC channel (parameterized by β). However, every T iterations, an LDPC bit-node is decimated (set to a fixed value based on the current

LLR). This encoding procedure for the codebook $\mathcal{C}(s^k)$ is described in Algorithm 1 assuming $s^k = 0^k$. For $s^k \neq 0^k$, the update for each check node in $\mathcal{P}_1$ is modified to include the appropriate s_k hard-decision message. Also, the decimated sequence u^n may not satisfy all the LDPC check constraints. In this case, successive encoding attempts often result in a valid codeword due to the randomization in Algorithm 1. In practice, we found that removing double edges and 4-cycles from the code essentially eliminated this problem at moderate block lengths. For example, see the results in Table III.1.

One variation in Algorithm 1 from [48] is the choice of the LDPC bit-node for decimation. In [48], bit-node with maximum bias over entire graph is selected, but we restrict the search for maximum biased bit to only left-most w sections of the bits that are not already decimated. We observed that this change increases the chances of encoding to a valid codeword.

Remark 68: The BPGD algorithm, when applied to *uncoupled* compound LDGM/ LDPC code, always failed to satisfy the LDPC check constraints and the spatial-coupling structure is required to overcome this problem. Thus, spatial coupling in compound codes not only helps to reduce the distortion, but allows the BPGD algorithm to encode to a valid codeword.

It is also possible to create a coupling structure in a circular fashion and decimate the bit-nodes in a fixed section as done in [83]. However, this leads to a *wave* of decimations from both ends and will result in a failure to encode to a codeword, as the constraints imposed by the LDPC check-nodes are unlikely to match when the two waves meet. From a physics point of view, this is akin to growing a crystal on a torus from a single seed. When the two growth interfaces meet, they are very unlikely to mesh nicely and form a pure crystal.

III.D.2 Iterative Quantization Algorithm

The encoding algorithm that is described below is introduced in the context of the BEQ problem as the iterative quantization algorithm [89]. In contrast to the BEQ problem, the additional challenge in WOM problem is for the quantized sequence to represent a given message s^k. This can be overcome by running the iterative quantization algorithm on the coset $\mathcal{C}(s^k)$.

The duality between the BEQ problem and the channel decoding over the BEC is further illustrated by the similarity between the iterative quantization algorithm and

Algorithm 2 Iterative Quantization Algorithm [89]

Input: Seq. $z^n \in \{0,1\}^n$, Msg. s^k, Comp. code $G(U,V)$.
Output: Seq. x^n in the $\mathcal{C}(s^k)$ that satisfies WOM const.

```
Associate z^n with LDGM check-nodes as in Fig. III.1.
Set all 0s in z^n to erasures ∗.
Set parities in P_1 to s^k.
while ∃ non-erasures in V do
  if ∃ non-erased u ∈ U such that only one of its neighbors v ∈ V is not erased
  then
    Pair (u, v).
    Erase u and v.
  else
    FAIL.
    break.
  end if
end while
if pairing did not FAIL then
  Set non-erased u ∈ U to 0.
  In the inverse order in which bit-nodes in U are erased, set them to satisfy the
  paired parity.
  Evaluate the codeword x^n from the graph G(U, V) and information bits u^m.
end if
```

the peeling decoder for the BEC. In the iterative quantization algorithm, degree-one bit-nodes are peeled off the graph rather than the degree-one check-nodes. Suppose $G(U,V)$ is a Tanner graph representation of the compound LDGM/LDPC, where U denotes the LDPC bit-nodes (nodes to be peeled off from the graph) and V denotes the union of check-nodes (analogous to the code-bits in peeling decoder). The details are presented in Algorithm 2.

Let z^n denote the state of the WOM system after first write. That is, 1's in z^n represent the WOM constraints. First, associate the sequence z^n with the LDGM check-nodes and set all 0's in z^n to erasures $*$. Now, LDPC bit-nodes which have a single non-erased check-node as its neighbor are paired with that unique check-node. Then, both those LDPC bit-nodes and the paired check-nodes are erased from the graph and this process is repeated. This is akin to the peeling decoder which assigns values to erased bit-nodes that have a degree-one check-node as its neighbor and then erases those bit-nodes. Once this pairing is done, the values of the LDPC bit-nodes are set to satisfy the paired check-node in the *reverse order that the LDPC bit-nodes were paired*.

Block length (n)	4-cycles	$1/2/3/4/\geq 5$
9000	yes	5/3/5/2/35
9000	no	21/12/5/3/9
27000	no	35/15/0/0/0
45000	no	40/9/0/0/1
63000	no	44/6/0/0/0
81000	no	50/0/0/0/0

Table III.1: Number of attempts for successful encoding for 50 codewords. Here, $d_v = 6$, $d_c = 3$, $d'_v = 3$, $d'_c = 6$, $(L, w) = (15, 3)$, $(\beta, T) = (0.65, 10)$.

For this algorithm to succeed in finding a desired codeword, there should be a sufficient number of degree-one LDPC bit-nodes to begin with. In fact, when this algorithm is used for the WOM problem with compound LDGM/LDPC codes and *regular* degrees, the process never succeeded. Therefore, it is required to design irregular degree distributions to have sufficient degree-one bit-nodes and achieve the optimal performance. However, this is not an issue in spatially-coupled codes since the boundary termination in these codes provides the necessary degree-one LDPC bit-nodes to get the encoding process started.

III.E NUMERICAL RESULTS

In Table III.1, we show the number of attempts required to encode 50 $\mathsf{Ber}(\frac{1}{2})$ source sequences in $\mathcal{C}(s^k)$ with the BPGD algorithm, with $s^k = 0^k$, over different block lengths. For example, at a block length of 9000, 21 sequences encoded in the first attempt and 9 sequences did not encode in four attempts. Without removing 4-cycles, only 5 sequences encoded at first and 35 failed after 4 attempts. At a block length of 81000, all 50 sequences encoded in the first attempt.

Consider a (d_v, d_c, d'_v, d'_c) compound LDGM/LDPC for the Wyner-Ziv problem with $k' = 0$ (i.e., empty $\mathcal{P}_2$). Then,

$$m = n\frac{d_c}{d_v}, \qquad\qquad k = m\frac{d'_v}{d'_c} = n\frac{d_c d'_v}{d_v d'_c}.$$

For the coding scheme described in Section III.C.2 for the Wyner-Ziv problem, this results in the optimal distortion D^* of

$$D^* = h^{-1}(1 - \text{code-rate of } \mathcal{C}')$$

LDGM/LDPC (d_v, d_c, d'_v, d'_c)	(L, w)	(D^*, p^*)	(D, p)
$(6, 3, 3, 6)$	(20,4)	(0.111,0.134)	(0.1174,0.122)
$(8, 4, 3, 6)$	(20,4)	(0.111,0.134)	(0.1149,0.120)
$(10, 5, 3, 6)$	(20,4)	(0.111,0.134)	(0.1139,0.122)

Table III.2: Thresholds for Wyner-Ziv problem with $n \approx 140000$, $\beta = 1.04$, $T = 10$.

LDGM/LDPC (d_v, d_c, d'_v, d'_c)	(L, w)	(δ^*, p^*)	(δ, p)
$(6, 3, 3, 6)$	(20,4)	(0.215,0.157)	(0.220,0.152)
$(8, 4, 3, 6)$	(20,4)	(0.215,0.157)	(0.223,0.151)
$(10, 5, 3, 6)$	(20,4)	(0.215,0.157)	(0.220,0.151)

Table III.3: Thresholds for Gelfand-Pinsker problem with $n \approx 140000$, $\beta = 0.65$, $T = 10$.

$$= h^{-1}\left(1 - \frac{m}{n}\right) = h^{-1}\left(1 - \frac{d_c}{d_v}\right).$$

From the capacity region in (III.1), the optimal p^* is given by

$$D^* * p^* = h^{-1}\left(h(D^*) + \frac{k}{n}\right),$$

which implies

$$p^* = \frac{h^{-1}\left(1 - \frac{d_c}{d_v} + \frac{d_c d'_v}{d_v d'_c}\right)}{1 - 2D^*}.$$

Consider the Gelfand-Pinsker problem with the compound LDGM/LDPC codes. For simplicity, we assume the parities in groups $\mathcal{P}_1$ (size k) and $\mathcal{P}_2$ (size k') have equal size and equal degrees. In the construction of the compound code, we ensure that all LDPC bit-nodes have equal degrees (except at the boundary in a spatially-coupled code) in both $\mathcal{C}$ and $\mathcal{C}(s^k)$. Thus, if the code $\mathcal{C}(s^k)$ has the degree profile (d_v, d_c, d'_v, d'_c), due to the equality assumptions about $\mathcal{P}_1$ and $\mathcal{P}_2$, the code $\mathcal{C}$ has the

degree profile $(d_v, d_c, d'_v/2, d'_c)$. With the above assumptions,

$$m = n\frac{d_c}{d_v}, \qquad\qquad k = k' = \frac{m}{2}\frac{d'_v}{d'_c} = \frac{n}{2}\frac{d_c d'_v}{d_v d'_c}.$$

For the coding scheme described in Section III.C.3 for the Gelfand-Pinsker problem, this results in

$$p^* = h^{-1}\left(1 - \left(\frac{d_c}{d_v} - \frac{1}{2}\frac{d_c d'_v}{d_v d'_c}\right)\right)$$
$$\delta^* = h^{-1}\left(1 - \left(\frac{d_c}{d_v} - \frac{d_c d'_v}{d_v d'_c}\right)\right).$$

We note that the rate loss in the spatially-coupled codes is not included when reporting the optimal thresholds. It is implicit that current constructions based on spatially-coupled codes suffer a rate loss of $O(w/L)$. While there has been some progress in mitigating this loss [91–93], minimizing this loss is an important open problem in the spatially-coupled codes. Also, for a fixed coupling window length w, the rate loss can be reduced arbitrarily by increasing L without changing the achievable threshold δ. As such, the thresholds shown are not sensitive to the parameter L.

Tables III.2 and III.3 provide the simulation results with spatially-coupled compound codes with message-passing algorithms. Encoding for the Wyner-Ziv problem is relatively easy, since this is performed using the codebook $\mathcal{SC}'$, which does not have LDPC check constraints. The reported distortion and the optimal thresholds correspond to the saturated section of the spatially-coupled system not effected by the boundary condition.

Consider a (d_v, d_c, d'_v, d'_c) compound LDGM/LDPC code as shown in for the second write of the noiseless 2-write WOM problem. We can assume $k' = 0$, since there is no need for error correction, and can have empty $\mathcal{P}_2$. Then,

$$m = n\frac{d_c}{d_v}, \qquad\qquad k = m\frac{d'_v}{d'_c} = n\frac{d_c d'_v}{d_v d'_c}.$$

For the second write, this gives a rate of

$$R = \frac{k}{n} = \frac{d_c d'_v}{d_v d'_c}.$$

LDGM/LDPC (d_v, d_c, d'_v, d'_c)	δ^*	δ $w=2$	δ $w=3$	δ $w=4$
$(3,3,3,6)$	0.500	0.477	0.492	0.494
$(3,3,4,6)$	0.333	0.294	0.324	0.326
$(3,3,5,6)$	0.167	0.095	0.156	0.158
$(4,4,3,6)$	0.500	0.461	0.491	0.492
$(4,4,4,6)$	0.333	0.278	0.323	0.325
$(4,4,5,6)$	0.167	0.086	0.155	0.159
$(5,5,3,6)$	0.500	0.436	0.488	0.491
$(5,5,4,6)$	0.333	0.260	0.320	0.324
$(5,5,5,6)$	0.167	0.079	0.154	0.159

Table III.4: Achievable threshold δ for the noiseless WOM system with spatially-coupled compound LDGM/LDPC codes with $L = 30$ and a single system blocklength of ≈ 24000.

Thus, from the capacity region in (III.4), the maximum normalized weight δ^* of the state sequence after first write below which we can successfully encode the second message is given by

$$\delta^* = 1 - R = 1 - \frac{d_c d'_v}{d_v d'_c}.$$

In Table III.4, we list the achievable thresholds δ for the spatially-coupled compound LDGM/LDPC codes. For the spatially-coupled codes, a chain length of $L = 30$ is used and the value of the threshold δ is shown for different coupling-window sizes. The blocklength of the single system is roughly of size 24000, which gives an effective blocklength for the coupled-system of about 720000. Such enormous blocklengths are required for the spatially-coupled codes to mitigate the rate loss and also operate close to capacity.

For the simulations, we have tested the encoding of 10 sequences with the iterative quantization algorithm, and the reported thresholds are the maximum values for which the majority of the 10 sequences are successfully encoded. A few observations are in order. From Table III.4, it is clear that one requires a coupling window lengths of at least $w = 3$ for the thresholds to have small gap to the optimal values. The decrease in the gap to the optimal threshold from changing $w = 3$ to $w = 4$ is minimal. To close the gap further, one needs to increase the blocklength of the single system beyond 24000.

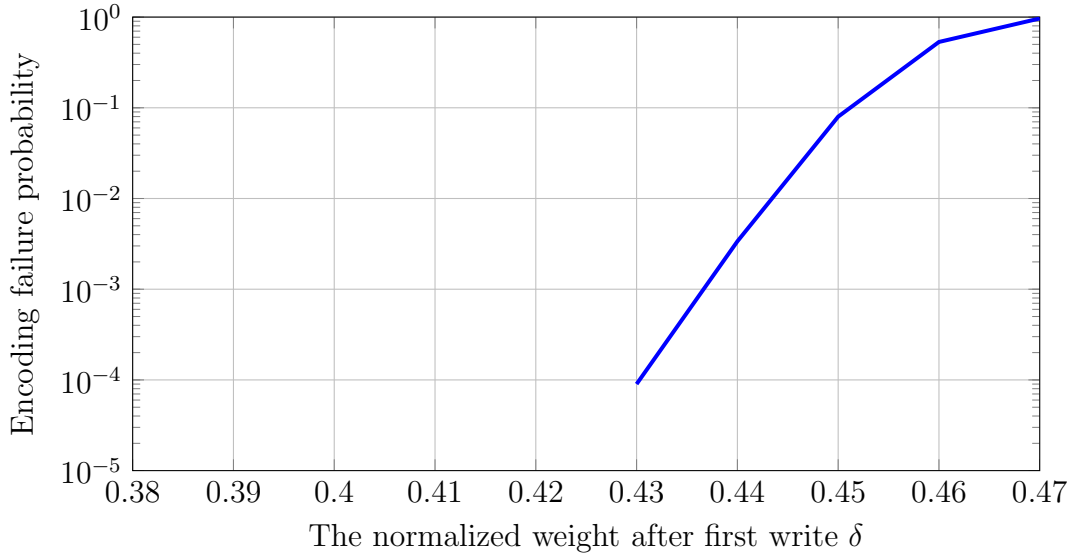

Figure III.3: Encoding failure probability for the second write as a function of the normalized weight after first write, for the spatially-coupled compound code with parameters $d_v = 3$, $d_c = 3$, $d'_v = 3$, $d'_c = 6$, $L = 30$, $w = 3$ and a single system block length of 1200. A total of 10^5 messages were attempted to encode, and no failures were observed for $\delta < 0.43$.

Next, we present the simulations for a smaller blocklength. In Fig. III.3, we show the encoding failure probability for the $(3, 3, 3, 6)$ spatially-coupled compound code with $(L, w) = (30, 3)$ and a single system blocklength of 1200, which gives an effective blocklength of 36000. A total of 10^5 messages were attempted to encode and no failures were observed for $\delta < 0.43$. These results appear to be better than the implementations based on polar codes. For example, in [73, Figure 2] for a polar code of length 16000 and rate 1/2, at $\delta = 0.42$ the encoding failure probability is more than 5×10^{-2}. However, note that our construction based on spatial coupling with parameters $(L, w) = (30, 3)$ suffers a rate loss of 10%. While the precise trade-offs between constructions based on spatially-coupled codes and polar codes are not completely clear, this is worth pursuing.

Now, lets consider the error-correcting WOM codes, where the message s^k has to be decoded from a noisy version of x^n corrupted by a binary symmetric channel with bit-flip probability p. For simplicity, we assume the parities in groups $\mathcal{P}_1$ (size k) and $\mathcal{P}_2$ (size k') have equal size and equal degrees. Recall that when decoding message s^k from a noisy version of x^n, the codebook of interest is $\mathcal{C}$ (where parities $\mathcal{P}_1$ are

LDGM/LDPC (d_v, d_c, d'_v, d'_c)	w	(δ^*, p^*)	(δ, p)
$(3, 3, 4, 6)$	3	$(0.333, 0.0615)$	$(0.321, 0.0585)$
$(3, 3, 4, 8)$	3	$(0.500, 0.0417)$	$(0.490, 0.0387)$
$(3, 3, 6, 8)$	4	$(0.250, 0.0724)$	$(0.239, 0.0684)$
$(4, 4, 4, 6)$	4	$(0.333, 0.0615)$	$(0.324, 0.0585)$
$(4, 4, 4, 8)$	4	$(0.500, 0.0417)$	$(0.492, 0.0387)$
$(4, 4, 6, 8)$	4	$(0.250, 0.0724)$	$(0.241, 0.0694)$

Table III.5: Achievable thresholds (δ, p) for the WOM system with read errors and spatially-coupled compound codes with $L = 30$ and a single system blocklength of ≈ 32000.

not present). In the construction of the compound code, we ensure that all LDPC bit-nodes have equal degrees (except at the boundary in a spatially-coupled code) in both $\mathcal{C}$ and $\mathcal{C}(s^k)$. Thus, if the code $\mathcal{C}(s^k)$ has the degree profile (d_v, d_c, d'_v, d'_c), due to the equality assumptions about $\mathcal{P}_1$ and $\mathcal{P}_2$, the code $\mathcal{C}$ has the degree profile $(d_v, d_c, d'_v/2, d'_c)$.

Consider a compound code $\mathcal{C}(s^k)$ with the degree profile (d_v, d_c, d'_v, d'_c). With the above assumptions,

$$m = n\frac{d_c}{d_v}, \qquad\qquad k = k' = \frac{m}{2}\frac{d'_v}{d'_c} = \frac{n}{2}\frac{d_c d'_v}{d_v d'_c}.$$

Thus, the rate of the second-write is given by

$$R = \frac{k}{n} = \frac{n}{2}\frac{d_c d'_v}{d_v d'_c}.$$

Since the degree profile of $\mathcal{C}$ is given by $(d_v, d_c, d'_v/2, d'_c)$, its code-rate is given by

$$\frac{m - k'}{n} = \frac{d_c}{d_v} - \frac{1}{2}\frac{d_c d'_v}{d_v d'_c} = \frac{d_c}{d_v}\left(1 - \frac{d'_v}{2d'_c}\right).$$

Therefore, the maximum channel parameter p^* below which the code $\mathcal{C}$ can correct errors with high probability is

$$p^* = h^{-1}(1 - \text{code-rate}) = h^{-1}\left(1 - \frac{d_c}{d_v}\left(1 - \frac{d'_v}{2d'_c}\right)\right).$$

From the rate region in (III.5), the maximum normalized weight δ^* of the state sequence after first write below which we can successfully encode the second message is given by

$$\delta^* = 1 - h(p^*) - R = \frac{d_c}{d_v} - \frac{d_c d'_v}{d_v d'_c}.$$

In Table III.5, the achievable thresholds (δ^*, p^*) are shown for the spatially-coupled compound codes. For these codes, a chain length of $L = 30$ is used and the thresholds are shown for different coupling window sizes. Again, we have not included the rate loss when reporting the optimal thresholds (δ^*, p^*). This results in a small loss in the optimal thresholds. The blocklength of the single system is roughly of size 30000.

III.F CONCLUSION

We have constructed spatially-coupled compound LDGM/ LDPC codes that achieve the capacity region of the binary instances of the Wyner-Ziv, Gelfand-Pinsker and, write-once memory (WOM) systems. The decoding and encoding is based on message-passing algorithms. For the Wyner-Ziv and Gelfand-Pinsker problems, the encoding is performed using the BPGD algorithm. Encoding in the compound codes is complicated since the LDPC bit-nodes need to satisfy additional constraints. The structure enforced by spatial-coupling seems to be crucial for the BPGD algorithm to encode to a codeword in compound codes.

For the WOM system, the focus was mostly on the 2-write WOM system. Encoding here is done by the iterative quantization algorithm from a reduction to the binary quantization problem. Finally, to the best of our knowledge, this construction appears to be the only non-polar coding scheme that can correct a constant fraction of errors in a WOM system with high probability.

CHAPTER IV

REED-MULLER CODES ACHIEVE CAPACITY ON ERASURE CHANNELS

IV.A INTRODUCTION

Since the introduction of channel capacity by Shannon in his seminal paper [1], theorists have been fascinated by the idea of constructing *structured* codes that achieve capacity. The advent of Turbo codes [2] and low-density parity-check (LDPC) codes [3, 5, 94] has made it possible to construct codes with low-complexity encoding and decoding that also achieve good performance near the Shannon limit. It was even proven that sequences of irregular LDPC codes can achieve capacity on the binary erasure channel (BEC) using low-complexity message-passing decoding [95]. For an arbitrary binary symmetric memoryless (BMS) channel, however, polar codes [6] were the first provably capacity-achieving codes with low-complexity encoding and decoding. More recently, spatially-coupled LDPC codes were also shown to achieve capacity universally over all BMS channels under low-complexity message-passing decoding [10, 18, 19, 90].

We consider the performance of sequences of binary linear codes transmitted over the BEC under maximum-a-posteriori (MAP) decoding. In particular, our primary technical result is the following.

Theorem: A sequence of linear codes achieves capacity on a memoryless erasure channel under MAP decoding if its blocklengths are strictly increasing, its code rates converge to some $r \in (0, 1)$, and the permutation group[1] of each code is doubly transitive.

Our analysis focuses primarily for the bit erasure rate under bit-MAP decoding but can be extended to the block erasure rate in some cases. One important consequence of this is that binary Reed-Muller codes achieve capacity on the BEC under block-MAP decoding.

The main result extends naturally to $\mathbb{F}_q$-linear codes transmitted over a q-ary erasure channel under symbol-MAP decoding. With this extension, one can show

[1]The permutation group of a linear code is the set of permutations on code bits under which the code is invariant.

that sequences of Generalized Reed-Muller codes [17,96] over $\mathbb{F}_q$ also achieve capacity under block-MAP decoding. For the class of affine-invariant $\mathbb{F}_q$-linear codes, which are precisely the codes whose permutation groups include a subgroup isomorphic to the affine linear group [97], one finds that these codes achieve capacity under symbol-MAP decoding. This follows from the fact that the affine linear group is doubly transitive. As it happens, this class also includes all extended primitive narrow-sense Bose-Chaudhuri-Hocquengham (BCH) codes [97]. Additionally, we show that sequences of extended primitive narrow-sense BCH codes over $\mathbb{F}_q$ achieve capacity under block-MAP decoding. To keep the presentation simple, we present proofs for the binary case and discuss the generalization to $\mathbb{F}_q$ in Section IV.E.4.

These results are rather surprising. Until the discovery of polar codes, it was commonly believed that codes with a simple deterministic structure might be unable to achieve capacity [98–100]. While polar codes might be considered a counterexample to this statement, they require a somewhat complicated design process that is heavily dependent on the channel. As such, their ability to achieve capacity appears somewhat unrelated to the inherent symmetry in the binary Hadamard transform. In contrast, the performance guarantees obtained here are a consequence only of linearity and the structure induced by the symmetry of the doubly-transitive permutation group.

Reed-Muller codes were introduced by Muller in [101] and, soon after, Reed proposed a majority logic decoder in [102]. A binary Reed-Muller code, parameterized by non-negative integers m and v, is a linear code of length 2^m and dimension $\binom{m}{0} + \cdots + \binom{m}{v}$. It is well known that the minimum distance of this code is 2^{m-v} [15–17]. Thus, it is impossible to simultaneously have a non-vanishing rate and a minimum distance that scales linearly with blocklength. As such, these codes cannot correct *all* patterns with a constant fraction of erasures. Until now, it was not clear whether or not these codes can correct *almost all* erasure patterns up to the capacity limit.

The idea that Reed-Muller codes might achieve capacity appears to be rather old. In a personal communication with Shu Lin, we learned that this possibility may have been discussed privately by Kasami, Lin, and Peterson as early as 1965. Later the idea was mentioned explicitly in a 1993 talk by Shu Lin, entitled "RM Codes are Not So Bad". To the best of the authors' knowledge, a 1994 paper by Dumer and Farrell contains the earliest printed discussion of this question [103]. In

that paper, they show that some sequences of BCH codes with rates approaching 1 have a vanishing gap to capacity on the BEC. They also suggest, as an open problem, the evaluation of a quantity which equals 1 if and only if Reed-Muller codes achieve capacity on the BEC. Since then, similar ideas have been discussed by a variety of authors [6, 14, 100, 104–107]. In particular, short Reed-Muller codes with erasures were investigated in [104, 105] and it was observed numerically that the block erasure rate is quite close to that of random codes. In [107], a modified construction of polar codes is analyzed and the results again suggest that Reed-Muller codes achieve capacity on the BEC. For rates approaching either 0 or 1 with sufficient speed, it has recently been shown by Abbe et al. that Reed-Muller codes can correct almost all erasure patterns up to the capacity limit[2] [14]. Beyond erasure channels, it is conjectured in [100] that the sequence of rate-1/2 self-dual Reed-Muller codes achieves capacity on the binary-input AWGN channel.

Even after 50 years of their discovery, Reed-Muller codes remain an active area of research in theoretical computer science and coding theory. The early work in [108–110] culminated in obtaining asymptotically tight bounds (fixed order v and asymptotic m) for their weight distribution [111]. Also, there is considerable interest in constructing low-complexity decoding algorithms [112–116]. Undoubtedly, interest in the coding theory community for these codes was rekindled by the tremendous success of polar codes and their close connection to Reed-Muller codes [6, 107, 117].

Due to their desirable structure, constructions based on these codes are used extensively in cryptography [104, 118–124]. Reed-Muller codes are also known for their *locality* [125]. Some of the earliest known constructions for locally correctable codes are based on these codes [126, 127]. Interestingly, local correctability of Reed-Muller codes is also a consequence of its permutation group being doubly transitive [128], a crucial requirement in our approach. However, a doubly transitive permutation group is not sufficient for local testability [129].

The central object in our analysis is the extrinsic information transfer (EXIT) function. EXIT charts were introduced by ten Brink [130] in the context of turbo decoding as a visual tool to understand iterative decoding. This work led to the area theorem for EXIT functions in [131] which was further developed in [132]. For a given input bit, the EXIT function is defined to be the conditional entropy of the

[2]It requires some effort to define precisely what capacity limit is for rates approaching 0 or 1. See [14, Definition 2.5] for details.

input bit given the outputs associated with all *other* input bits. The average EXIT function is formed by averaging all of the bit EXIT functions. We note that these functions are also instrumental in the design and analysis of LDPC codes [38].

An important property of EXIT functions is the EXIT area theorem, which says that the area under the average EXIT function equals the rate of the code. The value of the EXIT function at a particular erasure value is also directly related to the bit erasure probability under bit-MAP decoding. For a sequence of binary linear codes with rate r to be capacity achieving, the bit erasure probability, and therefore the average EXIT function, must converge to 0 for any erasure rate below $1 - r$. Since the areas under the average EXIT curves are fixed to r, the EXIT functions for these codes must also converge to 1 for any erasure rate above $1 - r$. Thus, the EXIT curves must exhibit a *sharp transition* from 0 to 1, and as a consequence of area theorem, this transition must occur at the erasure value of $1 - r$.

We investigate the threshold behavior of EXIT functions for certain binary linear codes via sharp thresholds for monotone boolean functions [133, 134]. The general method was pioneered by Margulis [135] and Russo [136]. Later, it was significantly generalized in [137] and [138]. This approach has been applied to many problems in theoretical computer science with remarkable success [139–141]. In the context of coding theory, Zémor introduced this approach in [142]. It was refined further in [143], and also extended to AWGN channels in [144]. For the BEC, [142, 143] show that the block erasure rate jumps sharply from 0 to 1 as the minimum distance of the code grows. However, this approach does not generalize directly to EXIT functions and, therefore, does not establish the location of the threshold. To show the threshold behavior for EXIT functions, we instead focus on symmetry [139] and require that the codes of interest have permutation groups that are doubly transitive.

After we completed this work [145], we discovered that the same approach was being pursued independently by Kudekar, Mondelli, Şaşoğlu, and Urbanke [146].

The chapter is organized as follows. Section IV.B includes necessary background on EXIT functions, permutation groups of linear codes, and capacity-achieving codes. Section IV.C deals with the threshold behavior of monotone boolean functions. In Section IV.D, as an application of the hitherto analysis, we show that Reed-Muller and extended primitive narrow-sense BCH codes achieve capacity. Finally, we provide extensions, open problems in Section IV.E, and concluding remarks in Section IV.F.

IV.B PRELIMINARIES

This chapter deals primarily with binary linear codes transmitted over erasure channels and bit-MAP decoding. In the following, all codes are understood to be proper binary linear codes with minimum distance at least 2, unless mentioned otherwise. Recall that a linear code is proper if no codeword position is 0 in all codewords. Let $\mathcal{C}$ denote an (N, K) binary linear code with length N and dimension K. The rate of this code is given by $r \triangleq K/N$. Denote the minimum distance of $\mathcal{C}$ by $d_{\min}$. We assume that a random codeword is chosen uniformly from this code and transmitted over a memoryless BEC. In the following subsections, we review several important definitions and properties related to this setup.

Notational convention:

- The natural numbers are denoted by $\mathbb{N} = \{1, 2, \ldots\}$.
- For $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$.
- We associate a binary sequence in $\{0,1\}^N$ with a subset of $[N]$ defined by the non-zero indices in the sequence. We use this equivalence between sets and binary sequences extensively. For example, a sequence 1001100 is identified by the subset $\{1, 4, 5\} \subseteq [7]$ and vice versa. Similarly, if 101110 is a codeword in $\mathcal{C}$, then we say $\{1, 3, 4, 5\} \in \mathcal{C}$.
- We say that a set A *covers* set B if $B \subseteq A$. Also, for sequences $\underline{a}, \underline{b} \in \{0,1\}^N$, we write $\underline{a} \leq \underline{b}$ if $a_i \leq b_i$ for $i \in [N]$. Equivalently, $\underline{a} \leq \underline{b}$ if the set associated with $\underline{b}$ covers the set associated with $\underline{a}$.
- For a set A, $\mathbb{1}_A(\cdot)$ denotes its indicator function. The random variable $\mathbb{1}_{\{\cdot\}}$ is an indicator of some event. For example, for random variables X and Y, $\mathbb{1}_{\{X \neq Y\}}$ indicates the event $X \neq Y$.
- For a vector $\underline{a} = (a_1, a_2, \ldots, a_N)$, the shorthand $\underline{a}_{\sim i}$ denotes

$$(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_N).$$

- 0^n and 1^n denote the all-zero and all-one sequences of length n, respectively.

- A memoryless BEC with erasure probability p is denoted by BEC(p). If the erasure probability is different for each bit, then we write BEC($\underline{p}$), where $\underline{p} = (p_1, \ldots, p_n)$ and p_i indicates the erasure probability of bit i.

- For a quantity θ with index n, we use either θ_n or $\theta^{(n)}$. Typically, we write $\theta^{(n)}$ when using θ_n may cause confusion with another quantity such as θ_i; in the latter case we write $\theta_i^{(n)}$.

- For a permutation $\pi \colon [N] \to [N]$ and $A \subseteq [N]$, $\pi(A)$ denotes the set $\{\pi(\ell) | \ell \in A\}$. For sequence $\underline{a} \in \{0,1\}^N$, $\pi(\underline{a})$ denotes the length-N sequence where the $\pi(i)$-th element is a_i.

- As is standard in information theory, $\mathrm{H}(\cdot)$ denotes the entropy of a discrete random variable and $\mathrm{H}(\cdot|\cdot)$ denotes the conditional entropy of a discrete random variable in bits.

- All logarithms in this chapter are natural unless the base is explicitly mentioned.

IV.B.1 Bit and Block Erasure Probability

The input and output alphabets of the BEC are denoted by $\mathcal{X} = \{0,1\}$ and $\mathcal{Y} = \{0,1,*\}$, respectively. Let $\underline{X} = (X_1, \ldots, X_N) \in \mathcal{X}^N$ be a uniform random codeword and $\underline{Y} = (Y_1, \ldots, Y_N) \in \mathcal{Y}^N$ be the received sequence obtained by transmitting $\underline{X}$ through a BEC(p). Here, our main interest is the bit-MAP decoder. But, we will also obtain some results for the block-MAP decoder indirectly based on our analysis of the bit-MAP decoder.

For linear codes and erasure channels, it is possible to recover the transmitted codeword if and only if the erasure pattern does not cover any codeword. To see this, fix an erasure pattern and observe that adding a codeword to the input sequence causes the output sequence to change if and only if the erasure pattern does not cover the codeword. Similarly, it is possible to recover bit i if and only if the erasure pattern does not cover any codeword where bit i is non-zero. Whenever bit i cannot be recovered uniquely, the symmetry of a linear code implies that set of codewords matching the unerased observations has an equal number of 0's and 1's in bit position i. In this case, the posterior marginal of bit i given the observations contains no information about bit i.

Let $D_i \colon \mathcal{Y}^N \to \mathcal{X} \cup \{*\}$ denote the bit-MAP decoder for bit i of $\mathcal{C}$. For a received sequence $\underline{Y}$, if X_i can be recovered uniquely, then $D_i(\underline{Y}) = X_i$. Otherwise, D_i declares an erasure and returns $*$. Let the erasure probability for bit $i \in [N]$ be

$$P_{b,i} \triangleq \Pr(D_i(\underline{Y}) \neq X_i),$$

and the average bit erasure probability be

$$P_b \triangleq \frac{1}{N} \sum_{i=1}^{N} P_{b,i}.$$

Whenever bit i can be recovered from a received sequence $\underline{Y} = \underline{y}$, $H(X_i|\underline{Y} = \underline{y}) = 0$. Otherwise, the uniform codeword assumption implies that the posterior marginal of bit i given the observations is $\Pr(X_i = x|\underline{Y} = \underline{y}) = \frac{1}{2}$ and $H(X_i|\underline{Y} = \underline{y}) = 1$. This immediately implies that

$$P_{b,i} = \mathrm{H}\left(X_i \mid \underline{Y}\right), \qquad\qquad P_b = \frac{1}{N} \sum_{i=1}^{N} \mathrm{H}\left(X_i \mid \underline{Y}\right).$$

Let $D \colon \mathcal{Y}^N \to \mathcal{X}^N \cup \{*\}$ denote the block-MAP decoder for $\mathcal{C}$. Given a received sequence $\underline{Y}$, the vector $D(\underline{Y})$ is equal to $\underline{X}$ whenever it is possible to uniquely recover $\underline{X}$ from $\underline{Y}$. Otherwise, D declares an erasure and returns $*$. Therefore, the block erasure probability is given by

$$P_B \triangleq \Pr(D(\underline{Y}) \neq \underline{X}).$$

Using the set equivalence

$$\{D(\underline{Y}) \neq \underline{X}\} = \bigcup_{i \in [N]} \{D_i(\underline{Y}) \neq X_i\},$$

it is easy to see that

$$P_{b,i} \leq P_B, \qquad\qquad P_b \leq P_B, \qquad\qquad P_B \leq N P_b. \tag{IV.1}$$

Also, if D declares an erasure, there will be at least $d_{\min}$ bits in erasure. Therefore,

$$d_{\min}\mathbb{1}_{\{D(\underline{Y})\neq\underline{X}\}} \leq \sum_{i\in[N]} \mathbb{1}_{\{D_i(\underline{Y})\neq X_i\}}.$$

Taking expectations on both sides gives a tighter bound on P_B in terms of P_b,

$$P_B \leq \frac{N}{d_{\min}} P_b. \tag{IV.2}$$

IV.B.2 MAP EXIT Functions

Again, let $\underline{X} = (X_1, \ldots, X_N)$ denote a uniformly selected codeword from $\mathcal{C}$ and $\underline{Y}$ be the sequence obtained from observing $\underline{X}$ with some positions erased. In this case, however, we assume X_i is transmitted over the BEC(p_i) channel. We refer to this as the BEC($\underline{p}$) channel where $\underline{p} = (p_1, \ldots, p_N)$ is the vector of channel erasure probabilities. While one typically evaluates all quantities of interest at $\underline{p} = (p, \ldots, p)$, such a parametrization provides a convenient mathematical framework for many derivations.

The vector EXIT function associated with bit i of $\mathcal{C}$ is defined by

$$h_i(\underline{p}) \triangleq \mathrm{H}\left(X_i|\underline{Y}_{\sim i}(\underline{p}_{\sim i})\right).$$

Also, the average vector EXIT function is defined by

$$h(\underline{p}) \triangleq \frac{1}{N}\sum_{i=1}^{N} h_i(\underline{p}).$$

Note that, while we define h_i as a function of $\underline{p}$ for uniformity, it does not depend on p_i. In terms of vector EXIT functions, the standard scalar EXIT functions $h(p)$ and $h_i(p)$ (for $i \in [N]$) are given by

$$h_i(p) \triangleq h_i(\underline{p})\Big|_{\underline{p}=(p,\ldots,p)}, \qquad h(p) \triangleq h(\underline{p})\Big|_{\underline{p}=(p,\ldots,p)}.$$

The bit erasure probabilities and the EXIT functions $h(p)$ and $h_i(p)$ have a close relationship. Observe that

$$\mathrm{H}\left(X_i \mid \underline{Y}\right) = \Pr(Y_i = *)\mathrm{H}\left(X_i|\underline{Y}_{\sim i}, Y_i = *\right) + \Pr(Y_i = X_i)\mathrm{H}\left(X_i|\underline{Y}_{\sim i}, Y_i = X_i\right)$$

$$= \Pr(Y_i = *)\mathrm{H}\left(X_i|\underline{Y}_{\sim i}\right).$$

Therefore,

$$P_{b,i}(p) = ph_i(p), \qquad P_b(p) = ph(p). \tag{IV.3}$$

We now state several well-known properties of these EXIT functions [131, 132], which play a crucial role in the subsequent analysis. It is worth noting that the original definition of EXIT charts in [131] focused on mutual information $I(\underline{X};\underline{Y})$ while later work on EXIT functions focused on the conditional entropy $H(\underline{X}|\underline{Y})$ [132]. In our setting, this difference results only in trivial remappings of all discussed quantities.

Proposition 69: For a code $\mathcal{C}$ on the BEC($\underline{p}$) channel, the EXIT function associated with bit i satisfies

$$h_i(\underline{p}) = \frac{\partial H(\underline{X}|\underline{Y}(\underline{p}))}{\partial p_i}.$$

For a parametrized path $\underline{p}(t) = (p_1(t), \ldots, p_n(t))$ defined for $t \in [0,1]$, where $p_i'(t)$ is continuous, one finds

$$\mathrm{H}\left(\underline{X}|\underline{Y}(\underline{1})\right) - \mathrm{H}\left(\underline{X}|\underline{Y}(\underline{0})\right) = \int_0^1 \left(\sum_{i=1}^{N} h_i(\underline{p}(t))p_i'(t)\right) dt.$$

Proof. This result is implied by the results of both [131] and [132]. For completeness, we repeat here the proof from [132, Theorem 2] using our notation.

For the first statement, we start by using chain rule of entropy to write

$$\mathrm{H}\left(\underline{X}|\underline{Y}(\underline{p})\right) = \mathrm{H}\left(X_i|\underline{Y}(\underline{p})\right) + \mathrm{H}\left(\underline{X}_{\sim i}|X_i, \underline{Y}(\underline{p})\right).$$

Then, we observe that

$$\mathrm{H}\left(\underline{X}_{\sim i}|X_i, \underline{Y}(\underline{p})\right) = \mathrm{H}\left(\underline{X}_{\sim i}|X_i, \underline{Y}_{\sim i}(\underline{p}_{\sim i})\right),$$

is independent of p_i. Since

$$\mathrm{H}\left(X_i|\underline{Y}(\underline{p})\right) = \Pr(Y_i = *)\mathrm{H}\left(X_i|\underline{Y}_{\sim i}(\underline{p}_{\sim i}), Y_i = *\right)$$

$$+ \Pr(Y_i = X_i)\mathrm{H}\left(X_i|\underline{Y}_{\sim i}(\underline{p}_{\sim i}), Y_i = X_i\right)$$
$$= p_i \mathrm{H}\left(X_i|\underline{Y}_{\sim i}(\underline{p}_{\sim i})\right),$$

we find that

$$\frac{\partial \mathrm{H}\left(X_i|\underline{Y}(\underline{p})\right)}{\partial p_i} = \mathrm{H}\left(X_i|\underline{Y}_{\sim i}(\underline{p}_{\sim i})\right) = h_i(\underline{p}).$$

The second statement now follows directly from vector calculus. □

The following sets characterize the EXIT functions h_i and we will refer to them throughout the chapter.

Definition 70: Consider a code $\mathcal{C}$ and the *indirect recovery* of X_i from the subvector $\underline{Y}_{\sim i}$ (i.e., the bit-MAP decoding of Y_i from $\underline{Y}$ when $Y_i = *$). For $i \in [N]$, the set of erasure patterns that prevent indirect recovery of X_i under bit-MAP decoding is given by

$$\Omega_i \triangleq \left\{A \subseteq [N]\backslash\{i\} \,|\, \exists B \subseteq [N]\backslash\{i\}, B \cup \{i\} \in \mathcal{C}, B \subseteq A\right\}.$$

For distinct $i, j \in [N]$, the set of erasure patterns where the j-th bit is *pivotal* for the indirect recovery of X_i is given by

$$\partial_j \Omega_i \triangleq \{A \subseteq [N]\backslash\{i\} \mid A\backslash\{j\} \notin \Omega_i, A \cup \{j\} \in \Omega_i\}.$$

These are erasure patterns where X_i can be recovered from $\underline{Y}_{\sim i}$ if and only if $Y_j \neq *$ (i.e., the j-th bit is not erased). Note that $\partial_j \Omega_i$ includes patterns from both Ω_i and Ω_i^c.

Intuitively, Ω_i is the set of all erasure patterns that cover some codeword whose i-th bit is 1. Also, since the minimum distance of $\mathcal{C}$ is at least 2 by assumption, the decoder can always recover bit i indirectly if no other bits are erased. Thus, Ω_i does not contain the empty set. For $j \in [N]\backslash i$, the set $\partial_j \Omega_i$ characterizes the boundary erasure patterns where flipping the erasure status of the j-th bit moves the pattern between Ω_i and Ω_i^c.

Proposition 71: For a code $\mathcal{C}$ on the BEC($\underline{p}$) channel, we have the following explicit expressions.

a) For bit i, the EXIT function is given by

$$h_i(\underline{p}) = \sum_{A \in \Omega_i} \prod_{\ell \in A} p_\ell \prod_{\ell \in A^c \backslash \{i\}} (1 - p_\ell).$$

b) For distinct i and j, the mixed partial derivative satisfies

$$\frac{\partial^2 H(\underline{X}|\underline{Y}(\underline{p}))}{\partial p_j \partial p_i} = \frac{\partial h_i(\underline{p})}{\partial p_j} = \sum_{A \in \partial_j \Omega_i} \prod_{\ell \in A} p_\ell \prod_{\ell \in A^c \backslash \{i\}} (1 - p_\ell).$$

Proof. For a), the definition of h_i implies

$$h_i(\underline{p}) = \mathrm{H}\left(X_i|\underline{Y}_{\sim i}(p_{\sim i})\right) = \sum_{\underline{y}_{\sim i} \in \mathcal{Y}^{N-1}} \Pr(\underline{Y}_{\sim i} = \underline{y}_{\sim i}) \mathrm{H}\left(X_i|\underline{Y}_{\sim i} = \underline{y}_{\sim i}\right).$$

Assume that all-zero codeword has been transmitted. Note that either $y_\ell = 0$ or $y_\ell = *$. Let $A \subseteq [N] \backslash \{i\}$ be the set of indices where $y_\ell = *$ so that

$$\Pr(\underline{Y}_{\sim i} = \underline{y}_{\sim i}) = \prod_{\ell \in A} p_\ell \prod_{\ell \in A^c \backslash \{i\}} (1 - p_\ell).$$

If $A \cup \{i\}$ covers a codeword in $\mathcal{C}$ whose i-th bit is non-zero, then bit-MAP decoder fails to decode bit i. Also, since the posterior probability of X_i given $\underline{Y}_{\sim i} = \underline{y}_{\sim i}$ is uniform, $H(X_i|\underline{Y}_{\sim i} = \underline{y}_{\sim i}) = 1$.

If $A \cup \{i\}$ does not cover any codeword in $\mathcal{C}$ with non-zero bit i, then the MAP estimate of X_i given $\underline{Y}_{\sim i} = \underline{y}_{\sim i}$ is equal to X_i and $H(X_i|\underline{Y}_{\sim i} = \underline{y}_{\sim i}) = 0$.

Thus, the EXIT function $h_i(\underline{p})$ is given by summing over the first set of erasure patterns where the entropy is 1. This set is precisely Ω_i, the set of all erasure patterns that cover a codeword whose i-th bit is non-zero.

For b), we evaluate the partial derivative using the explicit evaluation of $h_i(\underline{p})$ from part a). Suppose $A \in \Omega_i$. To simplify things, we handle the two groups separately.

If $A \cup \{j\} \in \Omega_i$ and $A \backslash \{j\} \in \Omega_i$, then we observe that

$$\sum_{B \in \{A \cup \{j\}, A \backslash \{j\}\}} \prod_{\ell \in B} p_\ell \prod_{\ell \in B^c \backslash \{i\}} (1 - p_\ell) = \prod_{\ell \in A \backslash \{j\}} p_\ell \prod_{\ell \in A^c \backslash \{i,j\}} (1 - p_\ell)$$

is independent of the variable p_j. Thus, its partial derivative with respect to p_j is zero.

On the other hand, if $A \cup \{j\} \in \Omega_i$ but $A \backslash \{j\} \notin \Omega_i$, then $j \in A$. In this case, the contribution of A to $h_i(\underline{p})$ can be written as

$$h_{i,A}(\underline{p}) = \prod_{\ell \in A} p_\ell \prod_{\ell \in A^c \backslash \{i\}} (1 - p_\ell).$$

Since $j \in A$, we find that

$$\frac{\partial h_{i,A}(\underline{p})}{\partial p_j} = \prod_{\ell \in A \backslash \{j\}} p_\ell \prod_{\ell \in A^c \backslash \{i\}} (1 - p_\ell) \tag{IV.4}$$

and, since the derivative is zero for patterns in the first group, we get

$$\frac{\partial h_i(\underline{p})}{\partial p_j} = \sum_{A \in \{B \in \Omega_i \,|\, B \backslash \{j\} \notin \Omega_i\}} \frac{\partial h_{i,A}(\underline{p})}{\partial p_j}. \tag{IV.5}$$

We can also rewrite (IV.4) as

$$\frac{\partial h_{i,A}(\underline{p})}{\partial p_j} = \sum_{B \in \{A \cup \{j\}, A \backslash \{j\}\}} \prod_{\ell \in B} p_\ell \prod_{\ell \in B^c \backslash \{i\}} (1 - p_\ell), \tag{IV.6}$$

where the effect of p_j is removed by summing over $A \cup \{j\}$ and $A \backslash \{j\}$. Substituting (IV.6) into (IV.5) gives the desired result because $\partial_j \Omega_i$ is equal to the union of $\{A \in \Omega_i \,|\, A \backslash \{j\} \notin \Omega_i\}$ and $\{A \notin \Omega_i \,|\, A \cup \{j\} \in \Omega_i\}$. □

The following proposition restates some known results in our notation. The area theorem, stated below as c), first appeared in [131, Theorem 1], and the explicit evaluation of $h_i(p)$, stated below in a), is a restatement of [38, Lemma 3.74(iv)].

Proposition 72: For a code $\mathcal{C}$ and transmission over a BEC, we have the following properties for the EXIT functions.

a) The EXIT function associated with bit i satisfies

$$h_i(p) = \sum_{A \in \Omega_i} p^{|A|} (1 - p)^{N - 1 - |A|}.$$

b) For $j \in [N]\backslash\{i\}$, the partial derivative satisfies

$$\frac{\partial h_i(\underline{p})}{\partial p_j}\Big|_{\underline{p}=(p,\ldots,p)} = \sum_{A \in \partial_j \Omega_i} p^{|A|}(1-p)^{N-1-|A|}.$$

c) The average EXIT function satisfies the *area theorem*

$$\int_0^1 h(p)dp = \frac{K}{N}.$$

Proof. The first two parts follow from Proposition 71. For the third part, we use Proposition 69(b) with the path $\underline{p}(t) = (t, \ldots, t)$. This gives

$$\mathrm{H}\left(\underline{X}|\underline{Y}(\underline{1})\right) - \mathrm{H}\left(\underline{X}|\underline{Y}(\underline{0})\right) = \int_0^1 \left(\sum_{i=1}^{N} h_i(t)\right) dt.$$

Also, $\mathrm{H}\left(\underline{X}|\underline{Y}(\underline{1})\right) = \mathrm{H}\left(\underline{X}\right) = K$ and $\mathrm{H}\left(\underline{X}|\underline{Y}(\underline{0})\right) = 0$. Combining these observations gives the desired result. □

Since the code $\mathcal{C}$ is proper by assumption, Ω_i is non-empty and, in particular, $[N]\backslash\{i\} \in \Omega_i$. Thus, h_i is not a constant function equal to 0 and $h_i(1) = 1$. Since the minimum distance of the code $\mathcal{C}$ is at least 2 by assumption, Ω_i does not contain the empty set. This implies that h_i is not a constant function equal to 1 and that $h_i(0) = 0$. As such, h_i is a non-constant polynomial. Also, h_i is non-decreasing because Proposition 72(b) implies that $dh_i/dp \geq 0$. It follows that h_i is strictly increasing because a non-constant non-decreasing polynomial must be strictly increasing.

Consequently, the EXIT functions $h_i(p)$, and therefore $h(p)$, are continuous, strictly increasing polynomial functions on $[0, 1]$ with $h(0) = h_i(0) = 0$ and $h(1) = h_i(1) = 1$.

The inverse function for the average EXIT function is therefore well-defined on $[0, 1]$. For $t \in [0, 1]$, let

$$p_t \triangleq h^{-1}(t) = \inf\{p \in [0, 1] \mid h(p) \geq t\}, \tag{IV.7}$$

and note that $h(p_t) = t$.

IV.B.3 Permutations of Linear Codes

Let S_N be the symmetric group on N elements. The permutation group of a code is defined as the subgroup of S_N whose group action on the bit ordering preserves the set of codewords [147, Section 1.6].

Definition 73: The permutation group $\mathcal{G}$ of a code $\mathcal{C}$ is defined to be

$$\mathcal{G} = \{\pi \in S_N \mid \pi(A) \in \mathcal{C} \text{ for all } A \in \mathcal{C}\}.$$

Interestingly, for binary linear codes, the permutation group is isomorphic to the group of weight-preserving linear transformations of the code [15, Section 8.5], [147, Section 7.9], [148].

Definition 74: Suppose $\mathcal{G}$ is a permutation group. Then,

a) $\mathcal{G}$ is *transitive* if, for any $i, j \in [N]$, there exists a permutation $\pi \in \mathcal{G}$ such that $\pi(i) = j$, and

b) $\mathcal{G}$ is *doubly transitive* if, for any distinct $i, j, k \in [N]$, there exists a $\pi \in \mathcal{G}$ such that $\pi(i) = i$ and $\pi(j) = k$.

Note that any non-trivial code (i.e., $0 < r < 1$) whose permutation group is transitive must be proper and have minimum distance at least two.

In the following, we explore some interesting symmetries of EXIT functions when the permutation group of the code is transitive or doubly transitive.

Proposition 75: Suppose the permutation group $\mathcal{G}$ of a code $\mathcal{C}$ is transitive. Then, for any $i \in [N]$,

$$h(p) = h_i(p) \qquad \text{for } 0 \leq p \leq 1.$$

Proof. Since $\mathcal{G}$ is transitive, for any $i, j \in [N]$, there exists a permutation π such that $\pi(i) = j$. This will allow us to show that there is a bijection between Ω_i and Ω_j induced by the action of π on the codeword indices. To do this, we first show that $A \in \Omega_i$ implies $\pi(A) \in \Omega_j$.

Since $A \in \Omega_i$, by definition, there exists $B \subseteq A$ such that $B \cup \{i\} \in \mathcal{C}$. Since $\pi \in \mathcal{G}$, $\pi(B \cup \{i\}) \in \mathcal{C}$. Also, $\pi(B \cup \{i\}) = \pi(B) \cup \{j\}$ and $\pi(B) \subseteq \pi(A)$. Consequently, $\pi(A) \in \Omega_j$.

Similarly, if $A \in \Omega_j$, then $\pi^{-1}(A) \in \Omega_i$. Thus, there is a bijection between Ω_i and Ω_j induced by π. This bijection also preserves the weight of the vectors in each set (i.e., $|A| = |\pi(A)|$).

Since Proposition 72(a) implies that $h_i(p)$ only depends on the weights of elements in Ω_i, it follows that $h_i(p) = h_j(p)$. This also implies that $h(p) = h_i(p)$ for all $0 \leq p \leq 1$. □

Proposition 76: Suppose that the permutation group $\mathcal{G}$ of a code $\mathcal{C}$ is doubly transitive. Then, for distinct $i, j, k \in [N]$, and any $0 \leq p \leq 1$,

$$\frac{\partial h_i(\underline{p})}{\partial p_j}\Big|_{\underline{p}=(p,\ldots,p)} = \frac{\partial h_i(\underline{p})}{\partial p_k}\Big|_{\underline{p}=(p,\ldots,p)}.$$

Proof. Since $\mathcal{G}$ is doubly transitive, there exists a permutation $\pi \in \mathcal{G}$ such that $\pi(i) = i$ and $\pi(j) = k$. Suppose $A \in \partial_j \Omega_i$. Then, by definition, either 1) $A \in \Omega_i$ and $A \backslash \{j\} \notin \Omega_i$ or 2) $A \cup \{j\} \in \Omega_i$ and $A \notin \Omega_i$. In either case, we claim that $\pi(A) \in \partial_k \Omega_i$. We prove this for the first case. The proof for the second case can be obtained verbatim by replacing A with $A \cup \{j\}$.

Suppose $A \in \Omega_i$ and $A \backslash \{j\} \notin \Omega_i$. Since $\pi \in \mathcal{G}$ and $\pi(i) = i$, $\pi(A) \in \Omega_i$. Also, $\pi(A \backslash \{j\}) \notin \Omega_i$; otherwise, $A \backslash \{j\} = \pi^{-1}(\pi(A \backslash \{j\})) \in \Omega_i$ gives a contradiction. Finally, $\pi(A \backslash \{j\}) = \pi(A) \backslash \{k\}$ implies that $\pi(A) \in \partial_k \Omega_i$. Similarly, one finds that $A \in \partial_k \Omega_i$ implies $\pi^{-1}(A) \in \partial_j \Omega_i$.

Since Proposition 72(b) implies that $\frac{\partial h_i}{\partial p_j}|_{\underline{p}=(p,\ldots,p)}$ only depends on the weights of elements in $\partial_j \Omega_i$ and $|A| = |\pi(A)|$, we obtain the desired result. □

IV.B.4 Capacity-Achieving Codes

Definition 77: Suppose $\{\mathcal{C}_n\}$ is a sequence of codes with rates $\{r_n\}$ where $r_n \to r$ for $r \in (0, 1)$.

a) $\{\mathcal{C}_n\}$ is said to be capacity achieving on the BEC under bit-MAP decoding, if for any $p \in [0, 1 - r)$, the average bit-erasure probabilities satisfy

$$\lim_{n \to \infty} P_b^{(n)}(p) = 0.$$

b) $\{\mathcal{C}_n\}$ is said to be capacity achieving on the BEC under block-MAP decoding, if

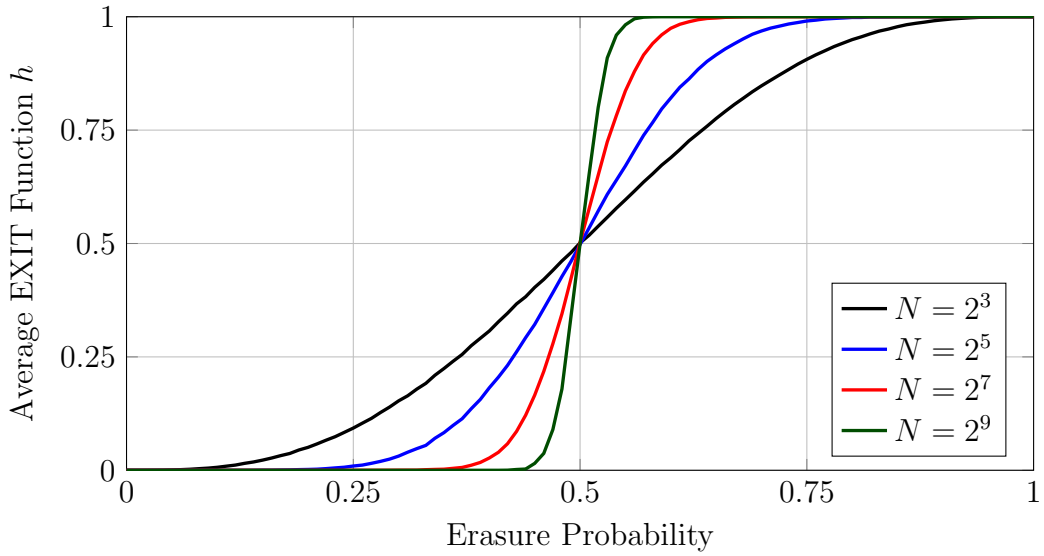

Figure IV.1: The average EXIT function of the rate-1/2 Reed-Muller code with blocklength N.

for any $p \in [0, 1-r)$, the block-erasure probabilities satisfy

$$\lim_{n\to\infty} P_B^{(n)}(p) = 0.$$

Note that in the definition above, we do not impose any constraints on the blocklength of the code $\mathcal{C}_n$.

The following proposition encapsulates the approach we use to show that a sequence of codes achieves capacity. It naturally bridges capacity-achieving codes, average EXIT functions, and the sharp transition framework presented in the next section, which allows one to show that the transition width[3] of certain functions converges to 0. The average EXIT functions of some rate-1/2 Reed-Muller codes are shown in Figure IV.1. Observe that as the blocklength increases, the transition width of the average EXIT function decreases. According to the following proposition, if this width converges to 0, then Reed-Muller codes achieve capacity on the BEC under bit-MAP decoding.

Proposition 78: Let $\{\mathcal{C}_n\}$ be a sequence of codes with rates $\{r_n\}$ where $r_n \to r$ for $r \in (0,1)$. Then, the following statements are equivalent.

[3]Defined as the width over which the function transitions from ε to $1-\varepsilon$.

S1: $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under bit-MAP decoding.

S2: The sequence of average EXIT functions satisfies

$$\lim_{n\to\infty} h^{(n)}(p) = \begin{cases} 0 & \text{if } 0 \le p < 1-r, \\ 1 & \text{if } 1-r < p \le 1. \end{cases}$$

S3: For any $0 < \varepsilon \le 1/2$,

$$\lim_{n\to\infty} \left(p_{1-\varepsilon}^{(n)} - p_{\varepsilon}^{(n)} \right) = 0,$$

where $p_t^{(n)}$ is the functional inverse of $h^{(n)}$ given by (IV.7).

Proof. See Section IV.G.1.

The equivalence between the first two statements is due to the close relationship between the bit erasure probability and the average EXIT function in (IV.3), while the equivalence between the last two statements is a consequence of the area theorem in Proposition 72(c). □

While the above result appears deceptively simple, our approach is successful largely because the transition point of the limiting EXIT function is known a priori due to the area theorem. Even though the sharp transition framework presented in the next section is widely applicable in theoretical computer science and allows one to deduce that the transition width of certain functions goes to 0, establishing the existence of a threshold and determining its precise location if it exists can be notoriously difficult [149–151].

IV.C SHARP THRESHOLDS FOR MONOTONE BOOLEAN FUNCTIONS VIA ISOPERIMETRIC INEQUALITIES

As seen in Proposition 78, the crucial step in showing that a sequence of codes achieves capacity is to prove that the average EXIT function transitions sharply from 0 to 1. From the explicit evaluation of h_i in Proposition 72(a), it is clear that the set Ω_i defines the behavior of h_i. Indeed, these sets play a crucial role in our analysis.

In this section, we treat the sets Ω_i and $\partial_j \Omega_i$ from Definition 70 as a set of sequences in $\{0,1\}^{N-1}$, since index i is not present in any of their elements. This occurs because $h_i(\underline{p})$ is not a function of p_i. To make this notion precise, we associate

$A \subseteq [N]\backslash\{i\}$ with $\Phi_i(A) \in \{0,1\}^{N-1}$, where bit ℓ of $\Phi_i(A)$ is given by

$$[\Phi_i(A)]_\ell \triangleq \begin{cases} \mathbb{1}_A(\ell) & \text{if } \ell < i, \\ \mathbb{1}_A(\ell+1) & \text{if } \ell \geq i. \end{cases}$$

Now, define

$$\begin{aligned} \Omega_i' &\triangleq \{\Phi_i(A) \in \{0,1\}^{N-1} \mid A \in \Omega_i\}, \\ \partial_j \Omega_i' &\triangleq \{\Phi_i(A) \in \{0,1\}^{N-1} \mid A \in \partial_j \Omega_i\}. \end{aligned} \tag{IV.8}$$

Whenever we treat Ω_i and $\partial_j \Omega_i$ as sequences of length $N-1$, we refer to them as Ω_i' and $\partial_j \Omega_i'$ to avoid confusion.

Consider the space $\{0,1\}^M$ with a measure μ_p such that

$$\mu_p(\Omega) = \sum_{\underline{x} \in \Omega} p^{|\underline{x}|}(1-p)^{M-|\underline{x}|}, \qquad \text{for } \Omega \subseteq \{0,1\}^M,$$

where the weight $|\underline{x}| = x_1 + \cdots + x_M$ is the number of 1's in $\underline{x}$. We note that $h_i(p) = \mu_p(\Omega_i')$ with $M = N-1$.

Recall that for $\underline{x}, \underline{y} \in \{0,1\}^M$, we write $\underline{x} \leq \underline{y}$ if $x_i \leq y_i$ for all $i \in [M]$.

Definition 79: A non-empty proper subset $\Omega \subset \{0,1\}^M$ is called *monotone* if $\underline{x} \in \Omega$ and $\underline{x} \leq \underline{y}$, then $\underline{y} \in \Omega$.

Remark 80: If the bit-MAP decoder cannot recover bit i from a received sequence, then it cannot recover bit i from any received sequence formed by adding additional erasures to the original received sequence. This implies that the set Ω_i' is monotone.

Monotone sets appear frequently in the theory of random graphs, satisfiability problems, etc. For a monotone set Ω, $\mu_p(\Omega)$ is a strictly increasing function of p. Often, the quantity $\mu_p(\Omega)$ exhibits a threshold type behavior, as a function of p, where it jumps quickly from 0 to 1. One technique that has been surprisingly effective in showing this behavior is based on deriving inequalities of the form

$$\frac{d\mu_p(\Omega)}{dp} \geq w\mu_p(\Omega)(1-\mu_p(\Omega)). \tag{IV.9}$$

If w is large, then the derivative of $\mu_p(\Omega)$ will be large when $\mu_p(\Omega)$ is not close to

either 0 or 1. In this case, $\mu_p(\Omega)$ must transition from 0 to 1 over a narrow range of p values.

One elegant way to obtain such inequalities is based on discrete isoperimetric inequalities [133,134]. First, let us define the function $g_\Omega \colon \{0,1\}^M \to \mathbb{N} \cup \{0\}$, which quantifies the boundary between Ω and Ω^c,

$$g_\Omega(\underline{x}) \triangleq \begin{cases} \left| \{\underline{y} \in \Omega^c \mid d_{\mathrm{H}}(\underline{x}, \underline{y}) = 1\} \right| & \text{if } \underline{x} \in \Omega, \\ 0 & \text{if } \underline{x} \notin \Omega, \end{cases} \tag{IV.10}$$

where d_{H} is the Hamming distance. Surprisingly, for a monotone set Ω, the derivative $d\mu_p(\Omega)/dp$ can be characterized exactly by g_Ω according to the Margulis-Russo Lemma [135, 136]:

$$\frac{d\mu_p(\Omega)}{dp} = \frac{1}{p} \int g_\Omega(\underline{x}) \mu_p(d\underline{x}).$$

Observe that $\mu_p(\Omega) + \mu_p(\Omega^c) = 1$ for any $0 \leq p \leq 1$. For a monotone set Ω, as we increase p, the probability from Ω^c flows to Ω. Intuitively, Margulis-Russo Lemma says that this flow depends only on the boundary between Ω and Ω^c. To obtain inequalities of type (IV.9), one approach is to find a lower bound on g_Ω that holds whenever it is non-zero [135, 136].

These techniques were introduced to coding by Tillich and Zémor to analyze the block error rate of linear codes under block-MAP decoding [142, 143]. In that case, the minimum non-zero value of g_Ω is proportional to the minimum distance of the code. Unfortunately, for the bit-MAP decoding problem we consider, the minimum non-zero g_Ω may be small (e.g., 1) even when the minimum distance of the code is arbitrarily large. This is discussed further in Section IV.E.1. To circumvent this, we discuss another approach, which requires a different formulation of the Margulis-Russo Lemma. We begin with a few definitions.

Definition 81: Let Ω be a monotone set and let

$$\partial_j \Omega \triangleq \left\{ \underline{x} \in \{0,1\}^M \mid \mathbb{1}_\Omega(\underline{x}) \neq \mathbb{1}_\Omega(\underline{x}^{(j)}) \right\},$$

where $\underline{x}^{(j)}$ is defined by $x_\ell^{(j)} = x_\ell$ for $\ell \neq j$ and $x_j^{(j)} = 1 - x_j$. Let the *influence of bit*

$j \in [M]$ be defined by

$$I_j^{(p)}(\Omega) \triangleq \mu_p\left(\partial_j \Omega\right)$$

and the *total influence* be defined by

$$I^{(p)}(\Omega) \triangleq \sum_{\ell=1}^{M} I_\ell^{(p)}(\Omega).$$

The Margulis-Russo Lemma can also be stated in terms of the total influence.

Theorem 82 ([133, Theorem 9.15])**:** Let Ω be a monotone set. Then,

$$\frac{d\mu_p(\Omega)}{dp} = I^{(p)}(\Omega).$$

Remark 83: Note that we have already seen Theorem 82 in the context of EXIT functions. When $M = N - 1$, it is easy to see from Proposition 72 that

$$h_i(p) = \mu_p(\Omega_i'), \qquad\qquad I_j^{(p)}(\Omega_i') = \frac{\partial h_i(\underline{p})}{\partial p_{j'}}\Big|_{\underline{p}=(p,\ldots,p)},$$

where

$$j' = \begin{cases} j & \text{if } j < i, \\ j+1 & \text{if } j \geq i. \end{cases} \tag{IV.11}$$

Therefore, Theorem 82 is equivalent to

$$\frac{dh_i(p)}{dp} = \sum_{j \in [N]\setminus\{i\}} \frac{\partial h_i(\underline{p})}{\partial p_j}\Big|_{\underline{p}=(p,\ldots,p)},$$

a straightforward result from vector calculus since h_i does not depend on p_i.

The advantage of using influences over g_Ω is that with "sufficient symmetry" in Ω, it is possible to show threshold phenomenon without any other knowledge about Ω. The following theorem illustrates the power of symmetry. Our proof hinges on this result.

Theorem 84: Let Ω be a monotone set and suppose that, for all $0 \leq p \leq 1$, the influences of all bits are equal $I_1^{(p)}(\Omega) = \cdots = I_M^{(p)}(\Omega)$.

a) Then, there exists a universal constant $C \geq 1$, which is independent of p, Ω, and M, such that

$$\frac{d\mu_p(\Omega)}{dp} \geq C(\log M)\mu_p(\Omega)(1-\mu_p(\Omega)),$$

for all $0 < p < 1$.

b) Consequently, for any $0 < \varepsilon \leq 1/2$,

$$p_{1-\varepsilon} - p_\varepsilon \leq \frac{2}{C}\frac{\log \frac{1-\varepsilon}{\varepsilon}}{\log M},$$

where $p_t = \inf\{p \in [0,1] \mid \mu_p(\Omega) \geq t\}$ is well-defined because $\mu_p(\Omega)$ is strictly increasing in p with $\mu_0(\Omega) = 0$ and $\mu_1(\Omega) = 1$.

Proof. See [138, 139, 152], [133, Section 9.6] for details.

In this form (i.e., by assuming all influences are equal), this result first appeared in [139]. However, this theorem can be seen as an immediate consequence of the earlier results in [153, Theorem 1], [138, Corollary 1.4]. The constant C was later improved in [152]. From the outline in [133, Exercise 9.8], one can verify this theorem for $C = 1$.

For the historical context, the study of influences for boolean functions was initiated in a 1987 technical report that led to [154]. Shortly after, [155] applied harmonic analysis to obtain some powerful general theorems about boolean functions. These results were subsequently generalized in [138, 153]. □

For the sets Ω_i', such a symmetry between influences is imposed by the doubly transitive property of the permutation group of the code according to Proposition 76.

Theorem 85: Let $\{\mathcal{C}_n\}$ be a sequence of codes where the blocklengths satisfy $N_n \to \infty$, the rates satisfy $r_n \to r$, and the permutation group $\mathcal{G}^{(n)}$ (of $\mathcal{C}_n$) is doubly transitive for each n. If $r \in (0,1)$, then $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under bit-MAP decoding.

Proof. Let the average EXIT function of $\mathcal{C}_n$ be $h^{(n)}$. The quantities N, $\mathcal{G}$, h, h_i, Ω_i', and p_t that appear in this proof are all indexed by n; we drop the index to avoid

cluttering. Fix some $i \in [N]$. Since $\mathcal{G}$ is transitive, from Proposition 75,

$$h(p) = h_i(p), \quad \text{for all } p \in [0,1].$$

Consider the sets Ω_i' from Definition 70 and (IV.8), and let $M = N - 1$. Observe that, from Proposition 72,

$$h_i(p) = \mu_p(\Omega_i'), \qquad\qquad I_j^{(p)}(\Omega_i') = \frac{\partial h_i(\underline{p})}{\partial p_{j'}}\Big|_{\underline{p}=(p,\ldots,p)},$$

where j' is given in (IV.11). Since $\mathcal{G}$ is doubly transitive, from Proposition 76,

$$I_j^{(p)}(\Omega_i') = I_k^{(p)}(\Omega_i') \quad \text{for all } j,k \in [N-1].$$

Using Theorem 84, we have

$$p_{1-\varepsilon} - p_\varepsilon \leq \frac{2}{C} \frac{\log \frac{1-\varepsilon}{\varepsilon}}{\log(N-1)},$$

where p_t is the functional inverse of h from (IV.7). Since $N \to \infty$ from the hypothesis,

$$\lim_{n\to\infty} (p_{1-\varepsilon} - p_\varepsilon) = 0.$$

Therefore, from Proposition 78, $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under bit-MAP decoding. □

We now focus on the block erasure probability. Recall from (IV.1) and (IV.2) that the block erasure probability satisfies the upper bounds

$$P_B \leq \frac{NP_b}{d_{\min}}, \qquad\qquad P_B \leq NP_b.$$

Thus, if $P_b \to 0$ with sufficient speed, then $P_B \to 0$ as well. Using (IV.9), one can derive the upper bound (see Lemma 97 in Section IV.G.2 for a proof)

$$\mu_\delta(\Omega) \leq \exp\left(-w[p_{1/2} - \delta]\right),$$

where $p_{1/2} \in [0,1]$ is defined uniquely by $\mu_{p_{1/2}}(\Omega) = 1/2$.

For $p < 1 - r$, the factor of $\log(N-1)$ in Theorem 85 determines the decay rate

of h with N, and consequently the decay rate of P_b. The following theorem shows that, if $d_{\min}$ satisfies $\log(d_{\min})/\log(N) \to 1$, then this decay rate is also sufficient to show that $P_B \to 0$.

Theorem 86: Let $\{\mathcal{C}_n\}$ be a sequence of codes where the blocklengths satisfy $N_n \to \infty$ and the rates satisfy $r_n \to r$ for $r \in (0,1)$. Suppose that the average EXIT function of $\mathcal{C}_n$ also satisfies, for $0 < p < 1$,

$$\frac{dh^{(n)}(p)}{dp} \geq C \log(N_n) h^{(n)}(p)(1 - h^{(n)}(p)),$$

where $C > 0$ is a constant independent of p and n. If the minimum distances $\{d_{\min}^{(n)}\}$ satisfy

$$\lim_{n\to\infty} \frac{\log d_{\min}^{(n)}}{\log N_n} = 1,$$

then $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under block-MAP decoding.

Proof. See Section IV.G.3. □

If $d_{\min}$ does not grow rapidly enough (e.g., sequences of Reed-Muller codes with rates $r_n \to r \in (0,1)$ have $d_{\min} = O(\sqrt{N}^{1+\delta})$ for any $\delta > 0$), then the previous theorem does not apply. Fortunately, it is possible to exploit symmetries, beyond the double transitivity of the permutation group, to obtain inequalities like (IV.9) that grow asymptotically faster than $\log(N)$ [156]. In particular, one obtains inequalities of type (IV.9), with factors of higher order than $\log(N)$, for all p except a neighborhood around 0 and 1 that vanishes as $N \to \infty$. The following theorem shows that this is sufficient to show that $P_B \to 0$ without imposing requirements on $d_{\min}$.

Theorem 87: Let $\{\mathcal{C}_n\}$ be a sequence of codes where the blocklengths satisfy $N_n \to \infty$ and the rates satisfy $r_n \to r$ for $r \in (0,1)$. Suppose that the average EXIT function of $\mathcal{C}_n$ also satisfies, for $a_n < p < b_n$,

$$\frac{dh^{(n)}(p)}{dp} \geq w_n \log(N_n) h^{(n)}(p)(1 - h^{(n)}(p)),$$

where $w_n \to \infty$, $a_n \to 0$, $b_n \to 1$ and $0 \leq a_n < b_n \leq 1$. Then, $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under block-MAP decoding.

Proof. See Section IV.G.4. □

IV.D APPLICATIONS

IV.D.1 Affine-Invariant Codes

Consider a code $\mathcal{C}$ of length $N = 2^m$ and the Galois field $\mathbb{F}_N$. Let $\Theta\colon [N] \to \mathbb{F}_N$ denote a bijection between the elements of the field and the code bits. Take a pair $\beta, \gamma \in \mathbb{F}_N$ with $\beta \neq 0$ and define $\pi_{\beta,\gamma} \in S_N$ such that

$$\pi_{\beta,\gamma}(\ell) = \Theta^{-1}(\beta\Theta(\ell) + \gamma).$$

Note that $\pi_{\beta,\gamma}$ is well-defined since Θ is bijective and $\beta \neq 0$, and observe that $\pi_{\beta_1,\gamma_1} \circ \pi_{\beta_2,\gamma_2} = \pi_{\beta_1\beta_2,\beta_1\gamma_2+\gamma_1}$. As such, the collection of permutations $\pi_{\beta,\gamma}$ forms a group. Now, the code $\mathcal{C}$ is called *affine-invariant* if its permutation group contains the subgroup

$$\{\pi_{\beta,\gamma} \in S_N \mid \beta, \gamma \in \mathbb{F}_N, \beta \neq 0\},$$

for some bijection Θ [147, Section 4.7].

Affine-invariant codes are of interest to us because their permutation groups are doubly transitive. To see this, consider distinct $i, j, k \in [N]$ and choose $\beta, \gamma \in \mathbb{F}_N$ where

$$\beta = \frac{\Theta(i) - \Theta(k)}{\Theta(i) - \Theta(j)}, \qquad \gamma = \Theta(i)\left(\frac{\Theta(k) - \Theta(j)}{\Theta(i) - \Theta(j)}\right),$$

and observe that $\pi_{\beta,\gamma}(i) = i$ and $\pi_{\beta,\gamma}(j) = k$.

Thus, by Theorem 85, a sequence of affine-invariant codes of increasing length, rates converging to $r \in (0, 1)$, achieve capacity on the BEC under bit-MAP decoding. Some examples of great interest include generalized Reed-Muller codes [17, Corollary 2.5.3] and extended primitive narrow-sense BCH codes [147, Theorem 5.1.9]. Below, we discuss Reed-Muller and BCH codes in more detail.

IV.D.2 Reed-Muller Codes

For integers v, m satisfying $0 \leq v \leq m$, the Reed-Muller code $\mathrm{RM}(v, m)$ is a binary linear code with length $N = 2^m$ and rate $r = 2^{-m}\left(\binom{m}{0} + \cdots + \binom{m}{v}\right)$. Although it is possible to describe these codes from the perspective of affine-invariance [17,

Corollary 2.5.3], below, we treat them as polynomial codes [157]. This provides a far more powerful insight to their structure [17, 158].

Consider the set of m variables, $x_1, \ldots, x_m$. For a monomial $x_1^{i_1} \cdots x_m^{i_m}$ in these variables, define its degree to be $i_1 + \cdots + i_m$. A polynomial in m variables is the linear combination (using coefficients from a field) of such monomials and the degree of a polynomial is defined to be the maximum degree of any monomial it contains. It is well-known that the set of all m-variable polynomials of degree at most v is a vector space over its field of coefficients. In this section, the coefficient field is the Galois field $\mathbb{F}_2$ and the vector space of interest is given by

$$P(m, v) = \text{span}\{x_1^{t_1} \ldots x_m^{t_m} \mid t_1 + \cdots + t_m \leq v, t_i \in \{0, 1\}\}.$$

For a polynomial $f \in P(m, v)$, $f(\underline{x}) \in \{0, 1\}$ denotes the evaluation of f at $\underline{x} \in \{0, 1\}^m$.

Let the elements of the vector space $\{0, 1\}^m$ over $\mathbb{F}_2$ be enumerated by $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_N$ with $\underline{e}_N = 0^m$. For any polynomial $f \in P(m, v)$, we can evaluate f at $\underline{e}_i$ for all $i \in [N]$. Then, the code $\text{RM}(v, m)$ is defined to be the set

$$\text{RM}(v, m) \triangleq \{(f(\underline{e}_1), \ldots, f(\underline{e}_N)) \mid f \in P(m, v)\}.$$

Lemma 88 ([96, Corollary 4])**:** The permutation group $\mathcal{G}$ of $\text{RM}(v, m)$ is doubly transitive.

Proof. Take any distinct $i, j, k \in [N]$. Below, we will produce a $\pi \in \mathcal{G}$ such that $\pi(i) = i$ and $\pi(j) = k$.

It is well known that for any vector space with two ordered bases $(\underline{u}_1, \ldots, \underline{u}_m)$ and $(\underline{u}'_1, \ldots, \underline{u}'_m)$, there exists an invertible $m \times m$ matrix T such that

$$\underline{u}_i = T\underline{u}'_i, \quad \text{for all } i \in [m].$$

Note that since i, j, k are distinct, $\underline{e}_j - \underline{e}_i \neq 0^m$ and $\underline{e}_k - \underline{e}_i \neq 0^m$. Therefore, there exists an invertible $m \times m$ binary matrix T such that $T(\underline{e}_j - \underline{e}_i) = \underline{e}_k - \underline{e}_i$.

For such a T, we construct $\pi \colon [N] \to [N]$ by defining $\pi(\ell) = \ell'$ for the unique ℓ' such that $\underline{e}_{\ell'} = T(\underline{e}_\ell - \underline{e}_i) + e_i$.

Note that $\pi \in S_N$ since T is invertible. Also, by construction, $\pi(i) = i$ and $\pi(j) = k$.

It remains to show that $\pi \in \mathcal{G}$. For this, consider a codeword in $\mathrm{RM}(v,m)$ given by $f \in P(m,v)$. It suffices to produce a $g \in P(m,v)$ such that $g(\underline{e}_{\pi(\ell)}) = f(\underline{e}_\ell)$ for all $\ell \in [N]$. Let

$$g(x_1,\ldots,x_m) = f(T^{-1}[x_1,\ldots,x_m]^{\mathrm{T}} - T^{-1}\underline{e}_i + \underline{e}_i),$$

and note that $\mathrm{degree}(f) = \mathrm{degree}(g)$, $g(\underline{e}_{\pi(\ell)}) = f(\underline{e}_\ell)$. Thus, we have the desired $g \in P(m,v)$. Hence, $\mathcal{G}$ is doubly transitive. □

There is also a sequence of $\{\mathrm{RM}(v_m,m)\}$ codes with increasing blocklengths and rates approaching any $r \in (0,1)$. To construct such a sequence, fix $r \in (0,1)$ and let $\{Z_i\}$ be an iid sequence of Bernoulli(1/2) random variables. Then, the rate of the $\mathrm{RM}(v_m,m)$ code is

$$\begin{aligned} r_m &= \frac{1}{2^m}\left(\binom{m}{0} + \cdots + \binom{m}{v_m}\right) \\ &= \Pr(Z_1 + \cdots + Z_m \le v_m) \\ &= \Pr\left(\frac{Z_1 - \frac{1}{2} + \cdots + Z_m - \frac{1}{2}}{\sqrt{m/4}} \le \frac{v_m - \frac{m}{2}}{\sqrt{m/4}}\right). \end{aligned}$$

Thus, by central limit theorem, if we choose

$$v_m = \max\left\{\left\lfloor \frac{m}{2} + \frac{\sqrt{m}}{2}Q^{-1}(1-r)\right\rfloor, 0\right\},$$

then the rate of $\mathrm{RM}(v_m,m)$ satisfies $r_m \to r$ as $m \to \infty$. Here,

$$Q(t) \triangleq \tfrac{1}{\sqrt{2\pi}}\int_t^\infty e^{-\tau^2/2}d\tau.$$

Theorem 89: For any $r \in (0,1)$, the sequence of codes $\{\mathrm{RM}(v_m,m)\}$ with

$$v_m = \max\left\{\left\lfloor \frac{m}{2} + \frac{\sqrt{m}}{2}Q^{-1}(1-r)\right\rfloor, 0\right\},$$

has rate $r_m \to r$ and is capacity achieving on the BEC under bit-MAP decoding.

Proof. This result follows as an immediate consequence of Lemma 88 and Theorem 85. □

We now analyze the block erasure probability of Reed-Muller codes. The minimum distance of Reed-Muller codes is too small to utilize Theorem 86. Thus, we use Theorem 87 instead.

For the code $\mathrm{RM}(v,m)$, consider the set Ω'_N from Definition 70 and (IV.8). Let $\mathcal{G}_N$ be the permutation group of Ω'_N defined by

$$\mathcal{G}_N \triangleq \{\pi \in S_{N-1} \mid \pi(\underline{a}) \in \Omega'_N \text{ for all } \underline{a} \in \Omega'_N\}.$$

Lemma 90: For the permutation group $\mathcal{G}_N$ defined above, there is a transitive subgroup isomorphic to $\mathrm{GL}(m,\mathbb{F}_2)$, the general linear group of degree m over the Galois field $\mathbb{F}_2$.

Proof. For a given $T \in \mathrm{GL}(m,\mathbb{F}_2)$, associate $\pi_T \in S_{N-1}$, where

$$\pi_T(\ell) = \ell', \quad \text{where } \underline{e}_{\ell'} = T\underline{e}_\ell.$$

Note that π_T is well-defined since T is invertible. Moreover, it is easy to check that $\pi_{T_1} \circ \pi_{T_2} = \pi_{T_1T_2}$ for $T_1, T_2 \in \mathrm{GL}(m,\mathbb{F}_2)$. As such, the collection of permutations

$$\mathcal{H} = \{\pi_T \in S_{N-1} \mid T \in \mathrm{GL}(m,\mathbb{F}_2)\}$$

is a subgroup of S_{N-1} isomorphic to $\mathrm{GL}(m,\mathbb{F}_2)$. Also, for $i,j \in [N-1]$, there exists $T \in \mathrm{GL}(m,\mathbb{F}_2)$ such that $\underline{e}_j = T\underline{e}_i$. For such a T, $\pi_T(i) = j$. Therefore, $\mathcal{H}$ is transitive.

It remains to show that $\mathcal{H} \subseteq \mathcal{G}_N$. For this, associate $\pi_T \in \mathcal{H}$ with $\pi'_T \in S_N$ where

$$\pi'_T(\ell) = \pi_T(\ell) \quad \text{for } \ell \in [N-1], \qquad\qquad \pi'_T(N) = N.$$

Also, it is easy to show that $\pi_T \in \mathcal{G}_1$ if $\pi'_T \in \mathcal{G}$, the permutation group of $\mathrm{RM}(v,m)$. To see that $\pi'_T \in \mathcal{G}$, consider a codeword given by $f \in P(m,v)$. It suffices to produce a $g \in P(m,v)$ where $g(\underline{e}_{\pi'_T(\ell)}) = f(\underline{e}_\ell)$ for $\ell \in [N]$. The desired g is given by $g(x_1,\ldots,x_m) = f(T^{-1}[x_1,\ldots,x_m]^{\mathrm{T}})$, by observing that $\mathrm{degree}(g) = \mathrm{degree}(f)$ and $g(\underline{e}_N) = f(T^{-1}0^m) = f(\underline{e}_N)$. □

Theorem 91: For any $r \in (0,1)$, the sequence of codes $\{\mathrm{RM}(v_m,m)\}$, with

$$v_m = \max\left\{\left\lfloor \frac{m}{2} + \frac{\sqrt{m}}{2}Q^{-1}(1-r)\right\rfloor, 0\right\},$$

has rate $r_m \to r$ and is capacity achieving on the BEC under block-MAP decoding.

Proof. Let the EXIT function associated with the last bit and the average EXIT function of the code $\mathrm{RM}(v_m, m)$ be h_N and h, respectively. Since the permutation group of $\mathrm{RM}(v_m, m)$ is transitive by Lemma 88, from Proposition 75, $h = h_N$. Moreover, by Lemma 90, $\mathcal{G}_N$ contains a transitive subgroup isomorphic to $\mathrm{GL}(m, \mathbb{F}_2)$.

Now, we can exploit the $\mathrm{GL}(m, \mathbb{F}_2)$ symmetry of Ω_N within the framework of [156]. In particular, [156, Theorem 1, Corollary 4.1] implies that there exists a universal constant $C > 0$, independent of m and p, such that

$$\frac{dh_N(p)}{dp} \geq C \log(\log N_m) \log(N_m) h_N(p)(1 - h_N(p)),$$

for $0 < a_m < p < b_m < 1$, where $N_m = 2^m$ and $a_m \to 0$, $b_m \to 1$ as $m \to \infty$. Since $h = h_N$, Theorem 87 implies that $\{\mathrm{RM}(v_m, m)\}$ is capacity achieving on the BEC under block-MAP decoding. □

From this, we see that the block erasure probability goes to 0 for $p < 1 - r$. For $p > 1 - r$, the average EXIT function $h(p)$ is bounded away from 0. Thus, Theorem 89 implies that the bit erasure probability $ph(p)$ is bounded away from 0 but not converging to 1. The block erasure probability does converge to 1, however. This follows from the result in [143] because the minimum distance of the code $\mathrm{RM}(v_m, m)$ goes to ∞ as $m \to \infty$.

IV.D.3 Bose-Chaudhuri-Hocquengham Codes

Let α be a primitive element of $\mathbb{F}_{2^m}$. Recall that a binary BCH code is *primitive* if its blocklength is of the form $2^m - 1$, and *narrow-sense* if the roots of its generator polynomial include consecutive powers of a primitive element starting from α. Here, we consider only primitive narrow-sense BCH codes and we follow closely the treatment of BCH codes in [147].

For integers v, m with $1 \leq v \leq 2^m - 1$, let $f(m, v)$ be the polynomial of lowest-degree over $\mathbb{F}_2$ that has the roots

$$\alpha, \alpha^2, \ldots, \alpha^v.$$

Then, $\mathrm{BCH}(v, m)$ is defined to be the binary cyclic code with the generator polynomial $f(m, v)$ and blocklength $N = 2^m - 1$. This is precisely the primitive narrow-sense

BCH code with blocklength N and designed distance $v+1$.

The dimension K of the cyclic code is determined by the degree of the generator polynomial according to [147, Theorem 4.2.1]

$$K = N - \text{degree}(f(m,v)).$$

Moreover, the minimum distance $d_{\min}$ of BCH(v,m) is at least $v+1$ [147, Theorem 5.1.1].

Since $\mathbb{F}_{2^m}$ is the splitting field of the polynomial $x^N - 1$ [147, Theorem 3.3.2], it is easy to see that $\text{degree}(f(m,N)) = N$. Also, since the size of the cyclotomic coset of any element α^i is at most m [147, Section 3.7], we have $\text{degree}(f(m,1)) \leq m$,

$$0 \leq \text{degree}(f(m,v+1)) - \text{degree}(f(m,v)) \leq m.$$

Thus, for any $r \in (0,1)$, one can choose $v_m \in [N]$ such that

$$N(1-r) \leq \text{degree}(f(m,v_m)) \leq N(1-r) + m.$$

Now, it is easy to see that $v_m \geq N(1-r)/m$ and the rate of the code BCH(v_m,m) will be in $[r - \frac{m}{N}, r]$.

Consider the length-2^m extended BCH code, eBCH(v,m), which is formed by adding a single parity bit to the code BCH(v,m) so that overall codeword parity is always even [147, Section 5.1]. The code eBCH(v,m) has the same dimension as BCH(v,m) and a minimum distance of at least $v+1$.

Thus, for any $r \in (0,1)$, there exists a sequence of codes $\{\text{eBCH}(v_m,m)\}$ with blocklengths $N_m = 2^m$, rates $r_m \to r$ and minimum distances

$$d_{\min}^{(m)} \geq 1 + v_m \geq 1 + \frac{N_m(1-r)}{m}. \tag{IV.12}$$

An important property of the extended BCH codes is that they are affine-invariant [147, Theorem 5.1.9]. Thus, Section IV.D.1 shows that their permutation group is doubly transitive. Therefore, we have the following theorem.

Theorem 92: For any $r \in (0,1)$, there is a sequence $\{v_m\}$ such that the code sequence $\{\text{eBCH}(v_m,m)\}$ has $r_m \to r$ and is capacity achieving on the BEC under bit-MAP decoding.

In the following, we discuss the block erasure probability of BCH codes. It is possible to characterize the permutation group of the code eBCH(v, m) precisely. According to [148, 159], except in sporadic cases, the permutation group of the code eBCH(v, m) is equal to the affine semi-linear group. Unfortunately, in the framework of [156], this group does not produce any factors beyond order $\log(N)$. This is not encouraging for the analysis of block erasure probability. This is in contrast with Reed-Muller codes where it was possible to exploit $\mathrm{GL}(m, \mathbb{F}_2)$ symmetry to analyze their block erasure probability. It is worth noting that the only primitive codes over a prime field, whose permutation group includes the general linear group of degree m, are variants of generalized Reed-Muller codes [158].

For BCH codes, however, the minimum distance is large enough to use Theorem 86. In fact, the minimum distance of the code eBCH(v_m, m) from (IV.12) satisfies

$$\lim_{m\to\infty} \frac{\log d_{\min}^{(m)}}{\log N_m} = 1. \tag{IV.13}$$

Since the permutation group of the code eBCH(v_m, m) is doubly transitive from affine-invariance, by Theorem 84 and the proof of Theorem 85, its average EXIT function satisfies the hypothesis of Theorem 86. Combining this observation with (IV.13) gives the following result.

Theorem 93: For any $r \in (0, 1)$, there is a sequence $\{v_m\}$ such that the code sequence $\{\mathrm{eBCH}(v_m, m)\}$ has $r_m \to r$ and is capacity achieving on the BEC under block-MAP decoding.

Corollary 94: For any $r \in (0, 1)$, there is a sequence $\{v_m\}$ such that the code sequence $\{\mathrm{BCH}(v_m, m)\}$ has $r_m \to r$ and is capacity achieving on the BEC under both bit-MAP and block-MAP decoding.

Proof. The code BCH(m, v) can be constructed from the code eBCH(m, v) simply by puncturing (i.e., erasing) the overall parity bit. This implies that their EXIT functions satisfy $h^{\mathrm{eBCH}}(p) \geq p h^{\mathrm{BCH}}(p)$. From this, we see that

$$h^{\mathrm{BCH}}(p) \leq \frac{1}{p} h^{\mathrm{eBCH}}(p).$$

Since puncturing single bit has an asymptotically negligible effect on the rate, the

statement of the corollary follows directly from Theorems 92 and 93. □

Remark 95: Corollary 94 shows that there are sequences of binary cyclic codes that achieve capacity on the BEC. As far as the authors know, this is the first proof that such a sequence exists [160].

IV.E DISCUSSION

IV.E.1 Comparison with the Work of Tillich and Zémor

Our initial attempts to prove a sharp threshold for EXIT functions focused on analyzing (IV.10) with $\Omega = \Omega_i$. In particular, our aim was to generalize [143] to EXIT functions by finding a lower bound on $g_{\Omega_i}(\underline{x})$ that holds uniformly over the boundary

$$\partial\Omega_i \triangleq \{\underline{x} \in \{0,1\}^N \,|\, g_{\Omega_i}(\underline{x}) > 0\}.$$

For code sequences where $d_{\min} \to \infty$ and the minimum distance of the dual code satisfies $d^{\perp}_{\min} \to \infty$, we expected that $\min_{\underline{x}\in\partial\Omega_i} g_{\Omega_i}(\underline{x})$ would grow without bound and, thus, that the EXIT function would have a sharp threshold. Unfortunately, this is not true. In fact, the ensemble of (j,k)-regular LDPC codes provides a counterexample. With high probability, their minimum distance grows linearly with N but one iteration of iterative decoding shows that the EXIT function is upper bounded by $(1-(1-p)^{k-1})^j$ for all p and N [38].

To understand this, first recall that a weight-d codeword in the dual code defines a subset of d code bits that sum to 0. If only one of the bits in this dual codeword is erased, then that bit can be recovered indirectly from the other bits. To see this in terms of the boundary, consider the indirect recovery of bit-i and assume that it is contained in a weight-d dual codeword with $d = d^{\perp}_{\min} \geq 3$. Let $\underline{x}$ be an erasure pattern where $d-2$ of the $d-1$ other bits in the dual codeword are received correctly and all other bits are erased. Then, $\underline{x} \in \Omega_i$ and bit-i cannot be recovered indirectly. Also, bit-i can be recovered indirectly if the erased bit (say bit j) in the dual codeword is revealed. Thus, $\underline{x}^{(j)} \notin \Omega_i$.

Now, let us consider $g_{\Omega_i}(\underline{x})$. If there is any other bit (say bit k) for which $\underline{x}^{(k)} \notin \Omega_i$, then the pattern of correctly received symbols in $\underline{x}^{(k)}$ (along with bit i) must cover a dual codeword. Since $\underline{x}^{(k)}$ contains exactly $d-1$ zero (i.e., unerased) symbols and the minimum dual distance is d, it follows that $\underline{x}^{(k)}$ must be a dual codeword. Due to linearity, one can add the two vectors to get $\underline{x}^{(j)} + \underline{x}^{(k)}$, which clearly has

weight 2. However, this contradicts the assumption that the minimum dual distance is $d^{\perp}_{\min} \geq 3$. Thus, we find that only bit j is pivotal for $\underline{x}$ and

$$\min_{\underline{x} \in \partial\Omega_i} g_{\Omega_i}(\underline{x}) = 1.$$

This shows that the method of [143] does not extend automatically to prove sharp thresholds for EXIT functions. While it is possible that there is a simple modification that overcomes this issue, we did not find it.

IV.E.2 Conditions of Theorem 85

One natural question is whether or not the conditions of Theorem 85 can be weakened. If the permutation groups of the codes in the sequence are not transitive, then different bits may have different EXIT functions with phase transitions at different values of p (e.g., if some of the bits are protected by a random code of one rate and other bits with a random code of a different rate).

Even if the permutation groups are transitive, things can still go wrong. Consider any sequence of codes with transitive permutation groups and increasing length. Let $\{d^{(n)}_{\min}\}$ be the sequence of minimum distances. Then, symmetry implies that the erasure rate of bit-MAP decoding is lower bounded by $p^{d^{(n)}_{\min}}$ for a BEC(p) (e.g., every code bit is covered by a codeword with weight $d_{\min}$). Thus, the sequence does not achieve capacity if $d^{(n)}_{\min}$ has a uniform upper bound. Based on duality, a similar argument holds if the sequence of minimum dual distances $\{d^{\perp(n)}_{\min}\}$ is upper bounded. Thus, to achieve capacity, a necessary condition is that $d^{(n)}_{\min} \to \infty$ and $d^{\perp(n)}_{\min} \to \infty$. Based on this observation, we make the following optimistic conjecture.

Conjecture 96: Let $\{\mathcal{C}_n\}$ be a sequence of binary linear codes where the blocklengths satisfy $N_n \to \infty$, the rates satisfy $r_n \to r$ for $r \in (0,1)$, and the permutation group of each code is transitive. If the sequence of minimum distances satisfies $d^{(n)}_{\min} \to \infty$ and the sequence of minimum dual distances satisfies $d^{\perp(n)}_{\min} \to \infty$, then the sequence achieves capacity on the BEC under bit-MAP decoding.

We call a code *reducible* if it can be written as the direct product of irreducible component codes of shorter length. If a code is reducible, then the minimum distance of each irreducible component is at least as large as the minimum distance of the overall code. Likewise, if the permutation group of a reducible code is transitive, then permutation group of each irreducible component code must also be transitive.

Moreover, transitivity implies that the EXIT function of each bit must equal both the EXIT function of the overall code and the EXIT function of any irreducible component code. Thus, the rate of the overall code and the rate of each irreducible component code must all be equal to the integral of their common EXIT function. This implies that, if the overall code satisfies the necessary conditions of the conjecture, then each of its irreducible component codes must also satisfy the necessary conditions. Thus, it is sufficient to resolve the conjecture for the case where there is a single irreducible component code.

IV.E.3 Beyond the Erasure Channel

Beyond the erasure channel, this work also has implications for the decoding of Reed-Muller codes transmitted over the binary symmetric channel. In particular, the results of [14, Theorem 1.8] show that an error pattern can be corrected by $\mathrm{RM}(m-(2t+2),m)$ whenever an erasure pattern with the same support can be corrected by $\mathrm{RM}(m-(t+1),m)$. Such error patterns can even be corrected efficiently [116].

Another interesting open question is whether or not one can extend this approach to binary-input memoryless symmetric channels via generalized EXIT (GEXIT) functions [44]. For this, some new ideas will certainly be required because the straightforward approach leads to the analysis of functions that are neither boolean nor monotonic.

It would also be very interesting to find boolean functions outside of coding theory where area theorems can be used to pinpoint sharp thresholds.

IV.E.4 $\mathbb{F}_q$-Linear Codes over the q-ary Erasure Channel

While our exposition focuses on binary linear codes over the BEC, it is easy to extend all results to $\mathbb{F}_q$-linear codes over the q-ary erasure channel.

First, the set Ω_i is redefined to be the set of erasure patterns that prevent indirect recovery of the symbol X_i. Importantly, Ω_i is still a set of binary sequences (equivalently, set of subsets of $[N]\backslash\{i\}$), and *not* a set of sequences over the alphabet $\{0,1,\ldots,q-1\}$. Note that, if indirect recovery is not possible, then the linearity of the code implies that the posterior marginal of symbol i given the extrinsic observations is $\Pr(X_i = x|\underline{Y}_{\sim i} = \underline{y}_{\sim i}) = 1/q$. Next, we rescale the logarithm in the entropy $\mathrm{H}(\cdot)$ to base q so that $H(X_i|\underline{Y}_{\sim i} = \underline{y}_{\sim i}) = 1$ when indirect recovery of X_i is not possible.

Thus, the sharp threshold framework for monotone boolean functions can be

applied without change. With these straightforward modifications, the results in Sections IV.B and IV.C hold true verbatim.

The concept of affine-invariance also extends naturally to $\mathbb{F}_q$-linear codes of length q^m over the Galois field $\mathbb{F}_q$. Similarly, affine-invariance implies that the permutation group is doubly transitive. Thus, sequences of affine-invariant $\mathbb{F}_q$-linear codes of increasing length, whose rates converge to $r \in (0,1)$, achieve capacity over the q-ary erasure channel under symbol-MAP decoding. The results for the block-MAP decoder also extend without change. Thus, one finds that Generalized Reed-Muller codes [17] and extended primitive narrow-sense BCH codes over $\mathbb{F}_q$ achieve capacity on the q-ary erasure channel under block-MAP decoding.

IV.E.5 Rates Converging to Zero

Consider a sequence of Reed-Muller codes $\{\mathrm{RM}(v_m, m)\}$ where the rate $r_m \to 0$ sufficiently fast. A key result of [14] is that Reed-Muller codes are capacity achieving in this scenario. That is, for any $\delta > 0$,

$$P_B^{(m)}(p_m) \to 0 \quad \text{for any } 0 \le p_m < 1 - (1+\delta) r_m.$$

Looking closely at [14, Corollary 5.1], it appears that $r_m = O(N_m^{-\kappa})$ for some $\kappa > 0$ is a necessary condition for this result, where the blocklength $N_m = 2^m$.

Let's analyze the bit erasure probability using our method. From the proof of Theorem 87, it is possible to deduce that $P_b^{(m)}(p_{\varepsilon_m}) \to 0$ if we choose $\varepsilon_m = o(1)$ such that $\log(1/\varepsilon_m) = o(\log(N_m))$.

We can also obtain a lower bound on p_{ε_m}. From the proof of Proposition 78, we gather that

$$p_{\varepsilon_m} \ge 1 - \frac{r_m}{1-\varepsilon_m} - \left(p_{1-\varepsilon_m} - p_{\varepsilon_m}\right).$$

From Theorem 84 and the proof of Theorem 85,

$$p_{1-\varepsilon_m} - p_{\varepsilon_m} \le \frac{2 \log \frac{1}{\varepsilon_m}}{\log(N_m - 1)},$$

which implies that

$$p_{\varepsilon_m} \ge 1 - \frac{r_m}{1-\varepsilon_m} - \frac{2 \log \frac{1}{\varepsilon_m}}{\log(N_m - 1)} = 1 - (1+\delta_m) r_m,$$

where

$$\delta_m = \frac{\varepsilon_m}{1-\varepsilon_m} + \frac{2\log\frac{1}{\varepsilon_m}}{r_m \log(N_m - 1)}.$$

Therefore,

$$P_b^{(m)}(p_m) \to 0 \quad \text{for any } 0 \le p_m < 1 - (1+\delta_m) r_m,$$

for any $\varepsilon_m = o(1)$ such that $\log(1/\varepsilon_m) = o(\log(N_m))$.

In order to obtain a capacity achieving result under bit-MAP decoding, we require that $\delta_m \to 0$. This can be guaranteed if $r_m \log(N_m) \to \infty$. Under this condition, we can choose $\varepsilon_m = 1/\log(r_m \log(N_m))$ so that

$$\varepsilon_m \to 0, \qquad \frac{\log\frac{1}{\varepsilon_m}}{\log(N_m)} \to 0, \qquad \delta_m \to 0.$$

Thus, under the condition $r_m \log(N_m) \to \infty$, the sequence $\mathrm{RM}(v_m, m)$ achieves capacity on the BEC under bit-MAP decoding.

For $r_m \to 0$, our results require $r_m \log(N_m) \to \infty$ while the results in [14, Corollary 5.1] require $r_m = O(N_m^{-\kappa})$ for some $\kappa > 0$. Thus, the results here apply to a different asymptotic rate regimes from [14].

IV.F CONCLUSION

We show that a sequence of binary linear codes achieves capacity if its block-lengths are strictly increasing, its code rates converge to some $r \in (0,1)$, and the permutation group of each code is doubly transitive. To do this, we use isoperimetric inequalities for monotone boolean functions to exploit the symmetry of the codes. This approach was successful largely because the transition point of the limiting EXIT function for the capacity-achieving codes is known a priori due to the area theorem. One remarkable aspect of this method is its simplicity. In particular, this approach does not rely on the precise structure of the code.

The main result extends naturally to $\mathbb{F}_q$-linear codes transmitted over a q-ary erasure channel under symbol-MAP decoding. The class of affine-invariant $\mathbb{F}_q$-linear codes also achieve capacity, since their permutation group is doubly transitive. Our results also show that Generalized Reed-Muller codes and extended primitive narrow-sense BCH codes achieve capacity on the q-ary erasure channel under block-MAP

decoding.

IV.G APPENDIX

IV.G.1 Proof of Proposition 78

$S1 \iff S2$: First, recall from (IV.3) that $P_b(p) = ph(p)$. From this, it follows that $S2 \implies S1$. Now, consider $S1 \implies S2$. The relation $P_b(p) = ph(p)$ together with $P_b^{(n)}(p) \to 0$ and $h^{(n)}(0) = 0$ implies

$$\lim_{n\to\infty} h^{(n)}(p) = 0 \quad \text{for } 0 \le p < 1 - r.$$

Now, we focus on the limit of $h^{(n)}(p)$ for $1-r < p \le 1$. Fix $q \in (1-r, 1]$ and choose n_0 large enough so that, for all $n > n_0$, we have $r_n > r - \varepsilon$ and $h^{(n)}(1 - r - \varepsilon) \le \varepsilon$. Such an n_0 exists because $r_n \to r$ and $h^{(n)}(p) \to 0$ for $0 \le p < 1 - r$. Since the function $h^{(n)}$ is increasing for all n, the EXIT area theorem (i.e., Proposition 72(c)) implies that, for all $n > n_0$, we have

$$\begin{aligned} r - \varepsilon < r_n &= \int_0^1 h^{(n)}(p)dp \\ &= \int_0^{1-r-\varepsilon} h^{(n)}(p)dp + \int_{1-r-\varepsilon}^{q} h^{(n)}(p)dp + \int_q^1 h^{(n)}(p)dp \\ &\le (1 - r - \varepsilon)\varepsilon + (q - (1 - r) + \varepsilon)h^{(n)}(q) + (1 - q). \end{aligned}$$

This implies

$$h^{(n)}(q) \ge \frac{q - (1 - r) - \varepsilon(2 - r - \varepsilon)}{q - (1 - r) + \varepsilon} \ge 1 - \frac{3\varepsilon}{q - (1 - r)}.$$

As such, $\lim_{n\to\infty} h^{(n)}(q) = 1$, for any $1 - r < q \le 1$.

$S2 \implies S3$: Since $p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)}$ is the width of the erasure probability interval over which $h^{(n)}$ transitions from ε to $1 - \varepsilon$, this follows immediately from $S2$.

$S3 \implies S2$: It suffices to show that, for any $\varepsilon \in (0, 1/2]$,

$$\lim_{n\to\infty} p_\varepsilon^{(n)} = \lim_{n\to\infty} p_{1-\varepsilon}^{(n)} = 1 - r.$$

From Proposition 72(c), we have

$$r_n = \int_0^1 h^{(n)}(\alpha)d\alpha \leq \varepsilon p_\varepsilon^{(n)} + \left(1 - p_\varepsilon^{(n)}\right),$$

which implies $p_\varepsilon^{(n)} \leq \frac{1-r_n}{1-\varepsilon}$. Similarly,

$$r_n = \int_0^1 h^{(n)}(\alpha)d\alpha \geq \left(1 - p_{1-\varepsilon}^{(n)}\right)(1-\varepsilon),$$

which implies $p_{1-\varepsilon}^{(n)} \geq \frac{1-r_n-\varepsilon}{1-\varepsilon}$.

Combining these gives

$$\frac{1-r_n-\varepsilon}{1-\varepsilon} + \left(p_\varepsilon^{(n)} - p_{1-\varepsilon}^{(n)}\right) \leq p_\varepsilon^{(n)} \leq \frac{1-r_n}{1-\varepsilon},$$
$$\frac{1-r_n-\varepsilon}{1-\varepsilon} \leq p_{1-\varepsilon}^{(n)} \leq \frac{1-r_n}{1-\varepsilon} + \left(p_{1-\varepsilon}^{(n)} - p_\varepsilon^{(n)}\right).$$

From the hypothesis,

$$\frac{1-r-\varepsilon}{1-\varepsilon} \leq \limsup_{n\to\infty} p_\varepsilon^{(n)} \leq \frac{1-r}{1-\varepsilon}, \quad \frac{1-r-\varepsilon}{1-\varepsilon} \leq \limsup_{n\to\infty} p_{1-\varepsilon}^{(n)} \leq \frac{1-r}{1-\varepsilon}.$$

Thus

$$\lim_{t\to 0}\limsup_{n\to\infty} p_t^{(n)} = \lim_{t\to 0}\limsup_{n\to\infty} p_{1-t}^{(n)} = 1-r.$$

But $p_t^{(n)}$ and $p_{1-t}^{(n)}$ are increasing and decreasing functions of t, respectively. This gives

$$\limsup_{n\to\infty} p_\varepsilon^{(n)} \geq \lim_{t\to 0}\limsup_{n\to\infty} p_t^{(n)} = 1-r, \quad \limsup_{n\to\infty} p_{1-\varepsilon}^{(n)} \leq \lim_{t\to 0}\limsup_{n\to\infty} p_{1-t}^{(n)} = 1-r.$$

Since $p_\varepsilon^{(n)} \leq p_{1-\varepsilon}^{(n)}$, we deduce that

$$\limsup_{n\to\infty} p_\varepsilon^{(n)} = \limsup_{n\to\infty} p_{1-\varepsilon}^{(n)} = 1-r.$$

Repeating this exercise with lim sup replaced by lim inf also gives the result $1-r$.

Thus, for any $\varepsilon \in (0, 1/2]$, we have

$$\lim_{n\to\infty} p_\varepsilon^{(n)} = \lim_{n\to\infty} p_{1-\varepsilon}^{(n)} = 1 - r.$$

IV.G.2 Proofs from Section IV.C

Lemma 97: Suppose $h\colon [0,1] \to [0,1]$ is a strictly increasing function with $h(0) = 0$ and $h(1) = 1$. Additionally, for $0 \le a < p < b \le 1$, let

$$\frac{dh(p)}{dp} \ge wh(p)(1 - h(p)).$$

If $p_t = h^{-1}(t)$, then for $0 < \varepsilon_1 \le \varepsilon_2 \le 1$,

$$p_{\varepsilon_2} - p_{\varepsilon_1} \le a + (1-b) + \frac{1}{w}\left[\log\frac{\varepsilon_2}{1-\varepsilon_2} + \log\frac{1-\varepsilon_1}{\varepsilon_1}\right]. \tag{IV.14}$$

Moreover, for $0 \le \delta \le p_{1/2}$,

$$h(\delta) \le \exp\left[-w\left([p_{1/2} - \delta] - [a + 1 - b]\right)\right].$$

Proof. Let $g(p) = \log\frac{h(p)}{1-h(p)}$ and observe that, for $a < p < b$, we have

$$\frac{dg(p)}{dp} = \frac{1}{h(p)(1-h(p))}\frac{dh(p)}{dp} \ge w.$$

Let $p_t = h^{-1}(t)$. We would like to obtain an upper bound on $p_{\varepsilon_2} - p_{\varepsilon_1}$ by integrating dg/dp.

If $a < p_{\varepsilon_1} \le p_{\varepsilon_2} < b$, then integrating dg/dp from p_{ε_1} to p_{ε_2} gives

$$w(p_{\varepsilon_2} - p_{\varepsilon_1}) \le \int_{p_{\varepsilon_1}}^{p_{\varepsilon_2}} \frac{dg}{dp}dp = \log\frac{\varepsilon_2}{1-\varepsilon_2} - \log\frac{\varepsilon_1}{1-\varepsilon_1},$$

which immediately shows (IV.14).

Suppose $p_{\varepsilon_1} \le a < p_{\varepsilon_2} < b$, and note that since g is increasing $\varepsilon_1 = g(p_{\varepsilon_1}) \le g(a)$. Then, integrating dg_1/dp from a to p_{ε_2} gives

$$w(p_{\varepsilon_2} - a) \le \int_a^{p_{\varepsilon_2}} \frac{dg}{dp}dp$$

$$
\begin{aligned}
&= \log \frac{\varepsilon_2}{1-\varepsilon_2} - \log \frac{h(a)}{1-h(a)} \\
&\leq \log \frac{\varepsilon_2}{1-\varepsilon_2} - \log \frac{\varepsilon_1}{1-\varepsilon_1}. \quad \text{(Since } \varepsilon_1 \leq h(a)\text{)}
\end{aligned}
$$

Using $p_{\varepsilon_2} - p_{\varepsilon_1} \leq a + (p_{\varepsilon_2} - a)$ with the above inequality gives (IV.14).

By considering other cases where p_{ε_1} and p_{ε_2} lie, it is straightforward to obtain (IV.14). Also, substituting $\varepsilon_2 = 1/2$ and $\varepsilon_1 = h(\delta)$ in (IV.14) gives the desired upper bound on $h(\delta)$. □

IV.G.3 Proof of Theorem 86

Let $p_t^{(n)}$ be the functional inverse of $h^{(n)}$ from (IV.7). Using Lemma 97 with $a_n = 0$ and $b_n = 1$ gives

$$
p_{1-\varepsilon}^{(n)} - p_{\varepsilon}^{(n)} \leq \frac{2 \log \frac{1-\varepsilon}{\varepsilon}}{C \log N_n}.
$$

By hypothesis, $N_n \to \infty$. Thus, for any $\varepsilon \in (0, 1/2]$, we have $p_{1-\varepsilon}^{(n)} - p_{\varepsilon}^{(n)} \to 0$. Using this, we apply statement $S2$ of Proposition 78 to see that $p_{1/2}^{(n)} \to 1 - r$.

Now, we can choose $\varepsilon_n = d_{\min}^{(n)}/(N_n \log N_n)$ and observe that

$$
\begin{aligned}
p_{1-\varepsilon_n}^{(n)} - p_{\varepsilon_n}^{(n)} &\leq \frac{2}{C \log N_n} \log \frac{1-\varepsilon_n}{\varepsilon_n} \\
&\leq \frac{2}{C \log N_n} \log \frac{N_n \log N_n}{d_{\min}^{(n)}} \\
&= \frac{2}{C} \frac{\log N_n + \log \log N_n - \log d_{\min}^{(n)}}{\log N_n}.
\end{aligned}
$$

By hypothesis, $\log d_{\min}^{(n)} / \log N_n \to 1$. Thus, $p_{1-\varepsilon_n}^{(n)} - p_{\varepsilon_n}^{(n)} \to 0$. Combining this with $p_{\varepsilon_n}^{(n)} \leq p_{1/2}^{(n)} \leq p_{1-\varepsilon_n}^{(n)}$ shows that $p_{\varepsilon_n}^{(n)} \to 1 - r$.

Also, from (IV.3),

$$
P_b^{(n)}(p_{\varepsilon_n}^{(n)}) = p_{\varepsilon_n}^{(n)} h^{(n)}(p_{\varepsilon_n}^{(n)}) \leq h^{(n)}(p_{\varepsilon_n}^{(n)}) = \varepsilon_n.
$$

Recall from (IV.2) that $P_B \leq N P_b / d_{\min}$. Hence, for any $p \in [0, 1-r)$, one finds that

$p_{\varepsilon_n}^{(n)} > p$ for sufficiently large n and thereafter

$$P_B^{(n)}(p) \leq \frac{N_n}{d_{\min}^{(n)}} P_b^{(n)}(p) \leq \frac{N_n}{d_{\min}^{(n)}} \varepsilon_n = \frac{1}{\log N_n} \to 0.$$

Thus, we conclude that $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under block-MAP decoding.

IV.G.4 Proof of Theorem 87

Let $p_t^{(n)}$ be the functional inverse of $h^{(n)}$ from (IV.7). From Lemma 97,

$$p_{1-\varepsilon}^{(n)} - p_{\varepsilon}^{(n)} \leq a_n + (1 - b_n) + \frac{2 \log \frac{1-\varepsilon}{\varepsilon}}{w_n \log N_n}.$$

By hypothesis, $a_n \to 0$, $1 - b_n \to 0$, and $w_n \log N_n \to \infty$. Thus, for any $\varepsilon \in (0, 1/2]$, we have $p_{1-\varepsilon}^{(n)} - p_{\varepsilon}^{(n)} \to 0$. Using this, we apply statement $S2$ of Proposition 78 to see that $p_{1/2}^{(n)} \to 1 - r$.

Now, we can choose $\varepsilon_n = 1/N_n^2$ and observe that

$$\begin{aligned} p_{1-\varepsilon_n}^{(n)} - p_{\varepsilon_n}^{(n)} &\leq a_n + (1 - b_n) + \frac{1}{w_n \log N_n} 2 \log \frac{1 - \varepsilon_n}{\varepsilon_n} \\ &\leq a_n + (1 - b_n) + \frac{1}{w_n \log N_n} 4 \log N_n \\ &= a_n + (1 - b_n) + \frac{4}{w_n}. \end{aligned}$$

Combining $p_{\varepsilon_n}^{(n)} \leq p_{1/2}^{(n)} \leq p_{1-\varepsilon_n}^{(n)}$ with $p_{1-\varepsilon_n}^{(n)} - p_{\varepsilon_n}^{(n)} \to 0$ shows that $p_{\varepsilon_n}^{(n)} \to 1 - r$.

Also, from (IV.3),

$$P_b^{(n)}(p_{\varepsilon_n}^{(n)}) = p_{\varepsilon_n}^{(n)} h^{(n)}(p_{\varepsilon_n}^{(n)}) \leq h^{(n)}(p_{\varepsilon_n}^{(n)}) = \varepsilon_n.$$

Recall from (IV.1) that $P_B \leq N P_b$. Hence, for any $p \in [0, 1 - r)$, one finds that $p_{\varepsilon_n}^{(n)} > p$ for sufficiently large n and thereafter

$$P_B^{(n)}(p) \leq N_n P_b^{(n)}(p) \leq N_n \varepsilon_n = N_n / N_n^2 \to 0.$$

Thus, we conclude that $\{\mathcal{C}_n\}$ is capacity achieving on the BEC under block-MAP decoding.

CHAPTER V

FIRST-PASSAGE TIME AND LARGE-DEVIATION ANALYSIS FOR ERASURE CHANNELS WITH MEMORY*

V.A INTRODUCTION

Contemporary communication systems must be designed to accommodate the various applications that compose today's digital landscape. In particular, mobile devices must meet the heterogeneous needs of various data flows in terms of delay tolerance and bandwidth requirements. On the Internet backbone, congestion is often prevented by over-provisioning. The large throughput and low latency of parallel optical lines provide a pragmatic solution that offers adequate network performance. This approach, combined with localized content distribution networks and edge throttling, is key in supporting delay-sensitive traffic over the Internet core. Unfortunately, a similar strategy cannot be applied to connect untethered devices, as wireless physical resources are limited and costly. The narrow usable spectrum and the broadcast nature of wireless environments limit the effective bandwidth of wireless access networks and, hence, demand the efficient management of available resources.

Here, we develop a mathematical framework that enables the optimal allocation of link resources for wireless systems in the context of delay-sensitive communication. Distinguishing features of the proposed methodology include the joint treatment of finite-state channels with memory and queueing behavior at the transmitter. The focus is on the first-passage time to an empty queue, and the methodology implicitly provides a distribution for the time it would take an additional packet to reach the head of the queue. This view is not only important for resource allocation and performance evaluation, it offers a foundation for choosing among possible routes and distinct interfaces. From an abstract perspective, we introduce a formulation where time-dependencies in channel states and decoding failures are captured meticulously.

In contrast to block-fading models, this formulation allows the seamless optimization of parameters such as code rate and block length. This is instrumental in better understanding how these parameters affect the overall performance of delay-sensitive wireless connections.

Several contributions on the interplay between decisions at the physical layer and overall performance at the link layer can be found in the literature [161–164]. Notable approaches include the outage capacity [165, 166], a probabilistic performance criterion based on the marginal distribution of channel blocks; the effective capacity [167, 168] which captures the decay rate in buffer occupancy at the transmitter; and finite block-length analyses of wireless connections [169, 170]. Physical resources can be optimized to reduce average delay by carefully selecting advantageous modulation schemes and coding strategies [171, 172]. Multi-objective problem formulations have also been explored. For instance, the optimal tradeoff between power and delay has received attention in the past [173]. The joint treatment of queueing and error-control coding has been examined by simultaneously considering the effective capacity of a link and the error exponent of a code family [174, 175]. Markov models have been successfully employed in the queueing analysis of communication links with automatic repeat request [176, 177]. Finally, powerful asymptotic techniques based on large deviations and heavy traffic limits have been developed to handle real-time traffic over unreliable links [178, 179].

This study differs from previous contributions in that it relates queueing behavior, error control coding and channel evolution without resorting to asymptotically long coding delays or rough approximations. Decoding performance at the receiver captures channel correlation within a block, while the queueing aspect of the problem is key in understanding the impact of time-dependencies among successive decoding attempts. Together, they provide an accurate assessment of overall system performance and lead to novel guidelines about efficient designs.

Furthermore, by focusing on the first-passage time to an empty queue [180], we are able to bypass the search for representative arrival processes. Rather, resource management can be performed adaptively based on current system conditions. Having a distribution for the hitting time to an empty buffer enables the computation of several pertinent performance criteria such as the probability of violating a completion deadline, the mean first-passage time to an empty queue, and Chernoff bounds. The proposed methodology is closely related to generating functions [181] and it

works well for reasonably small initial buffer sizes, which are typical of communication systems subject to stringent delay restrictions. On the other hand, under large buffers, this technique becomes somewhat cumbersome. In this latter case, analyzing the large deviations governing the evolution of the system offers a promising new direction to derive meaningful guidelines for resource allocation and the selection of system parameters. Indeed, the concentration of empirical measures can be used to gracefully adjust delay-sensitivity to the needs of real-time data flows by selecting the deviation threshold, i.e., the argument of the rate function [182]. Once a threshold is set, system parameters can be optimized according to this objective function and the resulting performance can be predicted accurately.

Throughout, we assume the availability of reliable acknowledgements using periodic feedback. We also assume that the transmitter and receiver share a common randomness, which permits the utilization of random binary codes. The remainder of this chapter is organized as follows. Section V.B presents the channel model and the random coding scheme. The queueing aspect of the problem is developed in Section V.C. A large deviations perspective on the mean transmission time and the average service rate is offered in Section V.D. The findings are supplemented by a discussion of pertinent criteria for performance evaluation, along with numerical examples. Concluding remarks and possible avenues of future research are exposed in Section V.G.

V.B SYSTEM MODEL

One physical aspect of wireless communication that we are particularly interested in is channel memory. From a queueing perspective, it is well known that correlation over time can drastically alter the stationary distribution of a queueing system [183, 184]. In a similar manner, channel memory can have a strong impact on overall performance, as it induces time-dependencies in the service process at the transmitter. This phenomenon is especially important for delay-sensitive applications that require the reliable, ordered delivery of data streams. A prime model class in dealing with such dependencies is composed of finite-state channels with memory [185–187]. System models derived from this class of channels are typically mathematically tractable, and they offer a natural mechanism to account for correlation over time. Moreover, insights acquired by studying erasure channels can often be translated to error channels or, at least, provide partial intuition about promising

solutions for the latter, more challenging scenarios.

Our discussion revolves around a communication paradigm where information bits flow from a source to a destination. The transmitter is assumed to possess a message of a certain length at the onset of the data transfer, and forward error correction is employed to shield content from potential symbol erasures. At the beginning of a transmission, the leading information bits stored at the source are grouped into a segment, and redundancy is added to this message using block encoding. The resulting codeword is then sent over a finite-state erasure channel with memory. Contingent upon the channel realization, the destination can either retrieve the data contained in the transmitted codeword or it declares a decoding failure. Successful transmissions are acknowledged and the corresponding bits are then discarded from the source buffer. Otherwise, the leading information bits remain in the queue. We emphasize that, in this framework, the original data sequence is guaranteed to be transferred unaltered. However, the completion time of the queue-emptying process is a random variable that depends on the coding/decoding strategy adopted and on the realization of the channel.

V.B.1 Channel Abstraction

As indicated above, we capture channel stochasticity and its impact on the communication link using a finite-state Markov process. Several pertinent communication scenarios can be modeled in this manner [188–190]. At a particular time instant, we assume that the channel can be in one of k states taking value in $\mathcal{C} = \{1, 2, \ldots, k\}$. State transitions over time form a Markov chain. We denote the corresponding transition probability matrix by

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{kk} \end{bmatrix}.$$

Entry b_{ij} in matrix $\mathbf{B}$ represents the conditional probability that, starting from state i, the channel transitions to state j. As such, $\mathbf{B}$ is a right stochastic matrix. When in state i, the transmitted symbol is erased with probability ε_i and, consequently, it is received correctly with probability $1 - \varepsilon_i$. For notational convenience, we impose a quality ordering on the channel states, i.e., $\varepsilon_i \geq \varepsilon_j$ whenever

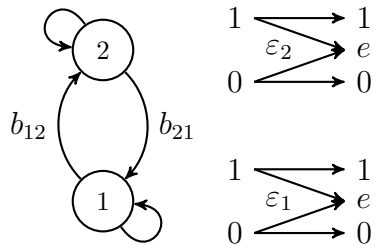

Figure V.1: Communication at the bit level takes place over a finite-state erasure channel with memory. While in state i, the probability of a bit erasure is ε_i. The evolution of the channel over time forms a Markov process.

$i < j$. We represent the state of the channel at time instant n by C_n. We note that $\{C_n\}$ is a first-order Markov process. A diagram illustrating the operation of the communication link for a two-state channel appears in Fig. V.1.

Assumption 98: Throughout, we hypothesize that the chain governing the finite-state channel is irreducible and aperiodic. We also assume that this Markov channel is non-trivial in that there exists a state $i \in \mathcal{C}$ such that $\varepsilon_i < 1$.

As we shall see, these conditions guarantee the existence of a random coding scheme for which the transmission process terminates in finite time, almost surely. These transmission schemes are the only ones of interest for our purpose. In that sense, Assumption 98 is introduced to prevent difficulties that arise from idiosyncratic, irrelevant scenarios.

V.B.2 Coding Scheme

The envisioned system employs forward error correction to counteract possible channel erasures. A codeword transmission attempt is initiated by selecting the leading K bits from the source buffer. Redundancy is then added to this data segment through the encoding process. A random coding scheme is adopted as a mathematically convenient abstraction to realistic implementations [161, 191]. To create each codeword transmission, a random binary parity check matrix of size $(N - K) \times N$ is generated. Every entry is selected uniformly over the binary alphabet, independently from other elements. The resulting codebook corresponds to the nullspace of this matrix. Such a coding scheme ensures that successful decoding of different codewords are conditionally independent given the channel states at the respective

transmission times. This will greatly simplify the ensuing analysis. We assume that maximum-likelihood decoding is performed at the receiver.

We emphasize that this mode of operation requires shared randomness at the source and the destination. Interestingly, this coding scheme is known to perform well for large block lengths; and it supports flexible rates of communication, any rate of the form K/N where $0 \leq K \leq N$ is admissible. These random codes have the additional property that the average probability of decoding failure depends only on the number of erasures caused by the channel and not on the specific locations of these erasures. Provided that e erasures have occurred during transmission, the probability of decoding failure can be evaluated explicitly,

$$P_{\mathrm{f}}(N-K, e) = 1 - \prod_{l=0}^{e-1} \left(1 - 2^{l-(N-K)}\right). \tag{V.1}$$

A proof for this statement is based on the equivalence between the linear independence of the e erased columns in the parity check matrix and the event of a successful decoding [191]. Throughout, $P_{\mathrm{f}}(p, e)$ denotes

$$P_{\mathrm{f}}(p, e) = \begin{cases} 1 - \prod_{l=0}^{e-1} \left(1 - 2^{l-p}\right) & \text{if } e \leq p \\ 1 & \text{if } p < e \leq N \end{cases} \tag{V.2}$$

which is the average probability of decoding failure under maximum likelihood of a codebook generated by using a random binary parity check matrix of size $p \times N$, for any $N \geq p$, when e erasures have occurred.

V.B.3 Distribution of Erasures

From the discussion above, we gather that the number of erasures suffered by a codeword plays a critical role in determining overall system performance, as it dictates the probability of decoding failure. This random variable thus warrants due attention. Let E denote the number of erasures occurring in a given packet transmission. Since the probability of decoding failure of a codeword depends only on the number of erasures, it suffices to consider probabilities of the form $\Pr(E = e, C_{N+1} = j | C_1 = i)$ to characterize the evolution of the system. Note that C_1 and C_{N+1} correspond to the channel state transitions across the first codeword transmission. We can describe this distribution in a compact form using matrix generating functions.

Define matrix $\mathbf{B}_x$ by

$$\mathbf{B}_x = \begin{bmatrix} b_{11}(1-\varepsilon_1+\varepsilon_1 x) & \cdots & b_{1k}(1-\varepsilon_1+\varepsilon_1 x) \\ b_{21}(1-\varepsilon_2+\varepsilon_2 x) & \cdots & b_{2k}(1-\varepsilon_2+\varepsilon_2 x) \\ \vdots & \ddots & \vdots \\ b_{k1}(1-\varepsilon_k+\varepsilon_k x) & \cdots & b_{kk}(1-\varepsilon_k+\varepsilon_k x) \end{bmatrix}.$$

Throughout, $[\![x^n]\!]$ denotes the linear operator that maps a polynomial in $\Re[x]$ to the coefficient of x^n. For $e \in \mathbb{N}_0$ and $i, j \in \mathcal{C}$, one can show that [181]

$$\Pr(E = e, C_{N+1} = j | C_1 = i) = [\![x^e]\!] \left[\mathbf{B}_x^N\right]_{i,j} \tag{V.3}$$

where, in this case, E denotes the number of erasures over an interval of length N. The probability that Markov process $\{C_n\}$ coincides with a specific sequence of states is equal to the probability of a certain path through the matching trellis. Moreover, at each point in time, the probability of observing an erasure only depends on the current state. Consequently, taking the Nth power of matrix $\mathbf{B}_x$ is an efficient way to compute the aggregate conditional probability of observing exactly e erasures, given an initial probability distribution and an end state. In other words, $\mathbf{B}_x^N$ offers a way to simultaneously sum all the relevant paths through the trellis. It is also possible to compute such probabilities through nested sums [192], but the ensuing equations rapidly become cumbersome for large values of N and Markov chains with sizable state spaces.

Given initial state i and for a fixed final state j, we can apply the total probability theorem to compute the probability of decoding failure,

$$\sum_{e=0}^{N} P_{\mathrm{f}}(N-K, e) \Pr\left(E = e, C_{N+1} = j | C_1 = i\right). \tag{V.4}$$

These conditional probabilities, along with the progression of the channel states, underlie the evolution of the queueing system.

Remark 99: As a side note, it is instructive to point out that, under Assumption 98, there exist values for N and K such that the probability of decoding success as a function of C_1 is not uniformly zero. In particular, if i is a channel state such that $\varepsilon_i < 1$, then for large enough N and $N-K$, the probability of decoding failure in

(V.4) will be less than one. Random codes for which the conditional probability of decoding success is not uniformly zero are termed non-trivial.

V.C QUEUEING MODEL

This section describes the queueing behavior of our system. First, we assume that the number of information bits present at the source at the beginning of the communication process is fixed and equal to ℓ. Given a code rate and block length, the source takes the leading K data bits and encodes the resulting segment into a codeword of length N using the scheme described in the preceding section. This codeword is then sent to the destination through N consecutive uses of the erasure channel. A service opportunity occurs every time the random code and channel realization jointly permit reliable decoding. We emphasize, again, that the destination is assumed to possess the ability to acknowledge the successful reception of codewords through instantaneous feedback. As such, the selected information bits remain in the transmit queue until a corresponding codeword is decoded faithfully at the destination. This data segment is immediately discarded from the buffer upon successful decoding of a packet.

In its simplest form, this scheme represents a variation of automatic repeat request (ARQ). We note that this mode of operation is somewhat naïve in that the information contained in failed decoding attempts is disregarded. A more astute implementation will seek to leverage past failures by performing joint decoding over all the observed messages pertaining to the current data segment. Incremental redundancy and hybrid automatic repeat request are valuable techniques that can improve performance [193–195]. Here, we discuss both ARQ and its hybrid variant, where partial information from failed transmission attempts is incorporated in the decoding process. Still, we focus largely on the rudimentary scheme because it admits a simpler, more elegant characterization while preserving the natural tradeoff between error protection and payload content. Overall, the proposed methodology yields pertinent results that help improve our understanding of delay-sensitive systems.

Our primary interest lies in the distribution of the time elapsed until the message originally contained in the source buffer becomes wholly available at the destination. To capture this quantity adequately, we need to examine the evolution of the queue. The length of the queue can be expressed in terms of the number of data segments awaiting transmission. If a queue initially contains ℓ information bits, then it will

require the successful reception of $m = \lceil \ell/K \rceil$ codewords until the last segment gets processed. The number of segments in the transmit buffer therefore becomes a measure of residual work until our objective is met, and it is intrinsically linked to the state of our communication system.

Codeword s denotes the block of transmitted bits during the time instants $sN + 1, \ldots, (s+1)N$, where $s \geq 0$. These codewords include both decoding successes and failures. For N fixed, we denote the size of the queue at the onset of codeword s by Q_s. We note that the state of the bit-erasure channel at the same time instant is C_{sN+1}. Thus at the onset of the first codeword transmission ($s = 0$), the size of the queue is Q_0 and the state of the bit-erasure channel is C_1. The rapid succession of symbols in the bit-erasure channel compared to events taking place in the queue produces the mismatch in indexing between Q_s and C_{sN+1}. Indeed, queue transitions are only possible at the completions of decoding attempts, which only occur after every N symbol transmissions. The resulting stochastic process $\{Q_s\}$ is a hidden Markov process, as it is determined partly by the evolution of the unobserved channel process $\{C_n\}$. While $\{Q_s\}$ alone does not possess the Markov property, it is possible to create an augmented process containing Q_s with this desirable attribute. The particulars of the procedure depend on whether one is considering the standard ARQ framework or its hybrid variant. We treat these two instances separately.

V.C.1 Automatic Repeat Request

As the title suggests, this section focuses exclusively on the scenario where the source and the destination employ ARQ to overcome channel erasures and, thereby, achieve reliable data transmission. In particular, the information contained in past decoding attempts is disregarded by the decoder when receiving the latest codeword. To build a suitable model, we consider the random vector $U_s = (C_{sN+1}, Q_s)$ composed of channel state and queue length. We wish to show that this vector contains all the relevant information to track the evolution of the system.

Theorem 100: The aggregate process $\{U_s\}_{s \geq 0}$ possesses the Markov property. That is, conditioned on $U_t = (i, q)$, the stochastic process $\{U_{s+t}\}_{s \geq 0}$ is independent of $U_0, \ldots, U_{t-1}$.

Proof. See Section V.H.1. □

Using the total probability theorem, we can write the transition probabilities of

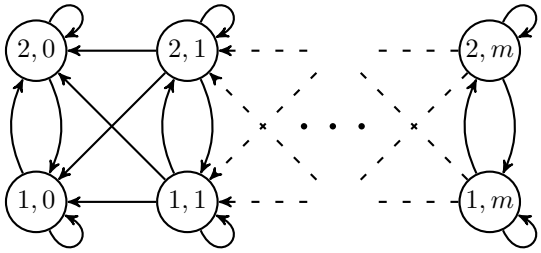

Figure V.2: This figure illustrates the progression of the queueing system for a service process that is governed by a two-state Markov erasure channel. System states, which are composed of queue lengths and channel states, are represented by circles. Admissible transitions are marked by the arrows.

$\{U_s\}$ as follows,

$$\begin{aligned}&\Pr\left(U_{s+1} = (j, q_{s+1}) | U_s = (i, q_s)\right) \\ &= \sum_{e=0}^{N} \Pr\left(Q_{s+1} = q_{s+1} | E = e, Q_s = q_s\right) \Pr\left(E = e, C_{(s+1)N+1} = j | C_{sN+1} = i\right) \quad \text{(V.5)}\end{aligned}$$

where $i, j \in \mathcal{C}$. For a non-empty queue, the first part of each summand corresponds to one of three possible cases,

$$\Pr\left(Q_{s+1} = q_{s+1} | E = e, Q_s = q_s\right) = \begin{cases} P_{\mathrm{f}}(N-K, e), & q_{s+1} = q_s \\ 1 - P_{\mathrm{f}}(N-K, e), & q_{s+1} = q_s - 1 \\ 0, & \text{otherwise.} \end{cases}$$

The probability of decoding failure $P_{\mathrm{f}}(\cdot, \cdot)$ appears in (V.1), while the conditional distribution of erasures within a block is given in (V.3). Thus, we have already developed the tools necessary to efficiently compute the value of every transition probability in (V.5). The evolution of the queueing system and its admissible transitions are depicted graphically in Fig. V.2.

The states $\{(\cdot, q)\}$ are collectively referred to as the qth level of the queue. The first-passage time to an empty buffer is therefore equivalent to the hitting time to level zero. Due to the repetitive structure of this augmented system, the hitting time to a lower level will play a key role in finding a tractable solution to the problem at hand.

An additional quantity of interest in the analysis of delay-sensitive systems is the mean service rate. To compute this quantity, it is convenient to analyze the service process $\{D_s\}$, where D_s indicates the potential of a successful decoding event at time s, $s \geq 0$. That is, $D_s = 1$ when a message can (or could) be decoded faithfully at the destination; and $D_s = 0$ otherwise. In words, the sequence $\{D_s\}$ indicates time instants at which blocks of information can be transferred successfully to the destination. As in the case of the queueing abstraction, the stochastic process $\{D_s\}$ forms a hidden Markov process which can be lifted to an augmented Markov process. Let $V_s = \big(C_{(s+1)N+1}, D_s\big)$ denote a random vector composed of the state of the erasure channel at the onset of block $s+1$, together with the indicator of a service opportunity during block s. As in Theorem 100, one can show that the stochastic process $\{V_s\}$ forms a Markov chain.

We note that the transition probabilities of $\{D_s\}$ are closely related to those of $\{Q_s\}$. Since there are no arrivals in our framework, the evolution of these processes are governed by

$$Q_{s+1} = (Q_s - D_s)^+ .$$

For convenience, we establish a succinct notation for the transition probabilities of our two augmented processes,

$$\begin{aligned} \kappa_{ij} &= \Pr(U_{s+1} = (j,q) | U_s = (i,q)) \\ &= \Pr(V_{s+1} = (j,0) | V_s = (i,d)) \\ \mu_{ij} &= \Pr(U_{s+1} = (j,q-1) | U_s = (i,q)) \\ &= \Pr(V_{s+1} = (j,1) | V_s = (i,d)) \end{aligned} \tag{V.6}$$

where $q \in \mathbb{N}$, $i, j \in \mathcal{C}$ and $d \in \{0,1\}$. These common definitions draw further attention to the close connection between $\{U_s\}$ and $\{V_s\}$.

In view of Remark 99 and for non-trivial codes, there exists $i \in \mathcal{C}$ such that $\mu_{ij} > 0$. This implies that the states associated with an empty buffer form the only closed communicating class and, as such, the remaining states are transient [180]. Since the number of states in the augmented chain is finite, this structure ensures that the task of emptying the transmit buffer is carried out in finite time, almost surely.

The symmetric decomposition of the queueing system into levels suggests an approach based on the quasi-birth-death structure of the chain. Suppose that the

buffer contains exactly m data segments at time zero, i.e., $Q_0 = m$. We can define the hitting time from level m to level q of the chain as

$$H_q = \inf\{s \geq 0 | Q_s = q\}, \tag{V.7}$$

where $0 \leq q < m$. That is, H_q designates the time instant at which the process $\{U_s\}$ first enters the qth level of the queue. We emphasize that, under the mild assumptions discussed above, H_q is almost surely finite. For consistency, we also define $H_m = 0$. Noting that Q_s is a non-increasing process, we can write the sojourn time at level q as

$$T_q = H_{q-1} - H_q,$$

where $0 < q \leq m$. That is, random variable T_q denotes the amount of time $\{U_s\}$ stays at level q before leaving for the subsequent lower level.

We are especially interested in H_0, the first-passage time to an empty queue. Taking advantage of the structure of the augmented Markov chain, we can fragment H_0 into a sum of elementary components. Specifically, the hitting time H_0 is equal to the sum of the sojourn times $T_1, \ldots, T_m$, i.e.,

$$H_0 = \sum_{q=1}^{m} T_q.$$

The sojourn times T_q and T_{q-1} are coupled through the channel state $C_{NH_{q-1}+1}$ and hence are not independent. However, since the codebooks over different codeword transmissions are independent, the sojourn times $T_1, \ldots, T_m$ are conditionally independent given the channel states $\left\{C_{NH_q+1}\right\}_{q=0}^{m}$. The sojourn times $T_1, \ldots, T_m$ are also conditionally identically distributed. That is,

$$\Pr\left(T_q = t, C_{NH_{q-1}+1} = j | C_{NH_q+1} = i\right)$$

is independent of q. A powerful means to compute the distribution of H_0 is to employ generating functions extended to matrices [181], exploiting the conditional independence and the identical distribution among the sojourn times $\{T_q\}$. This more intricate version of the generating function is necessary to keep track of the channel state entered after each downward queue transition. This method is described below.

Consider a reduced Markov chain composed of states $\{(i,0),(i,1)\}_{i=1}^{k}$, as shown

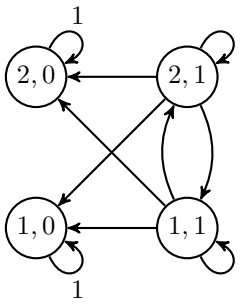

Figure V.3: This reduced Markov diagram represents one of the quasi-birth-death subcomponents of the queueing system. Starting from any distribution over these four states, it is possible to characterize the sojourn time T spent at level one. This is a key step in deriving the first-passage time to an empty buffer.

---

in Fig. V.3 for a Gilbert-Elliott channel. This reduced Markov chain represents one downward queue transition of the original system. Under proper state ordering, we can write the transition probability matrix for the reduced subsystem as

$$\mathbf{P} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{M} & \mathbf{K} \end{bmatrix}, \tag{V.8}$$

where we have implicitly defined matrices

$$\mathbf{M} = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1k} \\ \mu_{21} & \cdots & \mu_{2k} \\ \vdots & \ddots & \vdots \\ \mu_{k1} & \cdots & \mu_{kk} \end{bmatrix} \qquad \mathbf{K} = \begin{bmatrix} \kappa_{11} & \cdots & \kappa_{1k} \\ \kappa_{21} & \cdots & \kappa_{2k} \\ \vdots & \ddots & \vdots \\ \kappa_{k1} & \cdots & \kappa_{kk} \end{bmatrix}.$$

We emphasize that $\mathbf{P}$ is a stochastic matrix. As a consequence of the Perron-Frobenius theorem, we know that the spectral radius associated with $\mathbf{P}$ is one [196].

Define sojourn time T as the time spent at queue-level 1 of the reduced Markov chain. Mimicking our original notation, let Q_s denote the level of the queue (either 1 or 0) at the onset of codeword s and let $U_s = (C_{sN+1}, Q_s)$. Suppose the reduced Markov chain starts at queue-level 1, i.e. $Q_0 = 1$, then

$$T = \inf \{s \geq 0 | Q_s = 0\}.$$

The random variables $\{T_q\}_{q=1}^m$ and T have identical conditional distributions. That is, for any $1 \leq q \leq m$,

$$\Pr\left(T = t, C_{NT+1} = j | C_1 = i\right) = \Pr\left(T_q = t, C_{NH_{q-1}+1} = j | C_{NH_q+1} = i\right).$$

The distributions of the sojourn times $T_1, \ldots, T_m$ are important for determining the distribution of H_0. Thus, the above relation between $T_1, \ldots, T_m$ and T implies that the distribution of T is critical. Generating functions are an elegant way to characterize such distributions. Define matrix generating function $\mathbf{G}_T(z)$ entrywise by

$$\left[\mathbf{G}_T(z)\right]_{ij} = \mathrm{E}\left[z^T \mathbf{1}_{\{C_{NT+1}=j\}} | C_1 = i\right] \tag{V.9}$$

where $\mathbf{1}_{\{.\}}$ is the standard set indicator function.

Lemma 101: For the reduced subsystem associated with (V.8), the matrix generating function $\mathbf{G}_T(z)$ is equal to

$$\mathbf{G}_T(z) = \left(\mathbf{I} - \mathbf{K}z\right)^{-1} \mathbf{M}z. \tag{V.10}$$

Proof. The matrix generating function $\mathbf{G}_T(z)$ can be obtained by treating the entries of $\mathbf{P}$ as real polynomials in z, with

$$\mathbf{P}_z = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{M}z & \mathbf{K}z \end{bmatrix}.$$

Consider the two states $(i, 1)$ and (j, l), where $l = 0$ or $l = 1$. Their indices in the ordering associated with $\mathbf{P}$ are $k + i$ and $lk + j$, respectively. Recall that $[\![z^t]\!]$ denotes the operator that maps a polynomial in z to the coefficient of z^t. Suppose that, at time zero, the reduced system starts in state $(i, 1)$. After s transmissions, the reduced system will be in state $(j, 1)$ only when all the s transmissions result in decoding failures. Thus

$$\Pr\left(U_s = (j, 1) | U_0 = (i, 1)\right) = \left[\mathbf{K}^s\right]_{i,j}. \tag{V.11}$$

Similarly, the probability that the reduced system is in state $(j, 0)$ after s transmissions and having spent exactly t steps in queue-level 1, where $1 \leq t \leq s$, is given

by

$$\sum_{h=1}^{k} \Pr\left(U_s = (j,0), U_t = (j,0), U_{t-1} = (h,1) | U_0 = (i,1)\right).$$

Since the reduced system does not transition to a different state after reaching queue-level 0 (see Fig. V.3), this can be reduced to

$$\begin{aligned}
&\sum_{h=1}^{k} \Pr\left(U_s = (j,0), U_t = (j,0), U_{t-1} = (h,1) | U_0 = (i,1)\right) \\
&\qquad = \sum_{h=1}^{k} \Pr\left(U_t = (j,0), U_{t-1} = (h,1) | U_0 = (i,1)\right) \\
&\qquad = \left[\mathbf{K}^{t-1}\mathbf{M}\right]_{i,j}. \qquad \text{(V.12)}
\end{aligned}$$

Combining (V.11) and (V.12), the joint probability that the reduced system is in state (j,l) at time $s > 0$ and has spent exactly t steps at queue-level 1, where $1 \leq t \leq s$, can be expressed compactly as

$$\Pr(S_s = t, U_s = (j,l) | U_0 = (i,1)) = [\![z^t]\!] \left[\mathbf{P}_z^s\right]_{k+i,lk+j},$$

where S_s represents the total time spent at queue-level 1 over the interval from zero to instant s. Since T is a discrete random variable that is finite almost surely,

$$\begin{aligned}
\left[\mathbf{G}_T(z)\right]_{ij} &= \mathrm{E}\left[z^T \mathbf{1}_{\{C_{NT+1}=j\}} | C_1 = i\right] \\
&= \lim_{s\to\infty} \sum_{t=0}^{s} \Pr\left(T = t, C_{Nt+1} = j | C_1 = i\right) z^t \\
&= \lim_{s\to\infty} \sum_{t=0}^{s} \Pr\left(S_s = t, U_s = (j,0) | U_0 = (i,1)\right) z^t \\
&= \lim_{s\to\infty} \sum_{t=0}^{s} \left([\![z^t]\!] \left[\mathbf{P}_z^s\right]_{k+i,j}\right) z^t \\
&= \lim_{s\to\infty} \left[\mathbf{P}_z^s\right]_{k+i,j}.
\end{aligned}$$

Therefore the generating matrix $\mathbf{G}_T(z)$ can be obtained as

$$\begin{aligned}
\mathbf{G}_T(z) &= \lim_{s\to\infty} \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{P}_z^s \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \\
&= \lim_{s\to\infty} \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \sum_{t=1}^{s} \mathbf{K}^{t-1}\mathbf{M}z^t & \mathbf{M}^s z^s \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \\
&= \lim_{s\to\infty} \sum_{t=1}^{s} \mathbf{K}^{t-1}\mathbf{M}z^t \\
&= (\mathbf{I} - \mathbf{K}z)^{-1}\,\mathbf{M}z.
\end{aligned}$$

The above equation holds for all $|z| < \varrho(\mathbf{K})^{-1}$, where $\varrho(\cdot)$ denotes the spectral radius of its matrix argument. □

V.C.2 Hybrid Automatic Repeat Request

Hybrid ARQ is a mechanism that seeks to incorporate the partial information contained in failed transmissions into the subsequent decoding attempts of the same data segment. In this sense, it differs significantly from ARQ only when the initial decoding of a data segment fails. For finite-state erasure channels with memory, the evolution of a hybrid ARQ system can be characterized completely, although in a somewhat cumbersome manner. To implement hybrid ARQ with random codes, we must modify our coding strategy slightly.

Herein, we focus on hybrid schemes with finite depths. That is, the transmitter-receiver pair has a predetermined number of tries to successfully transmit a data segment. Our favored implementation relies on puncturing random codes. In a way analogous to our previous approach, we generate a codebook by creating a random binary parity check matrix of size $(aN - K) \times aN$, where a is the depth of the hybrid ARQ scheme. Again, the entries are selected uniformly from the binary alphabet and the codebook is equal to the nullspace of this matrix. The hybrid ARQ scheme progresses as follows. First, an information segment is mapped to a codeword of length aN. During the initial transmission, the leading N symbols of this codeword are sent over the erasure channel. Upon completion of this phase, the destination tries to recover the original data segment. When decoding fails, the next N symbols are sent and the aggregate message is run through a maximum-likelihood decoder. This process continues, communicating N symbols at a time, until the message is

successfully decoded at the destination or the total number of attempts reaches its limit.

Since untransmitted symbols can be classified as erasures for the purpose of decoding, we can leverage (V.2) in assessing the probabilities of decoding failure at the destination. That is, when s codeword chunks are present at the destination, out of which a total of e symbols are erased, the probability of decoding failure can be written as

$$P_{\mathrm{f}}(aN-K, e+(a-s)N) = 1 - \prod_{i=0}^{e+(a-s)N-1} \left(1 - 2^{i-(aN-K)}\right). \tag{V.13}$$

Comparing this expression for $s = 1$ and $a > 1$ to (V.1), we gather that the probability of decoding failure after receiving one chunk of length N for the hybrid ARQ scheme differs from the probability of failure in standard ARQ. Indeed, there is a slight penalty for the initial transmission resulting from using a random code tailored to hybrid ARQ. The following proposition establishes a uniform bound on the loss in performance associated with the hybrid scheme.

Proposition 102: Suppose that p and e are fixed, positive integers. The function of n defined by

$$P_{\mathrm{f}}(p+n, e+n) = \begin{cases} 1 - \prod_{l=0}^{n+e-1} \left(1 - 2^{l-p-n}\right) & \text{if } e \leq p \\ 1 & \text{if } e > p \end{cases}$$

is monotone increasing. Furthermore, the difference between this function and $P_{\mathrm{f}}(p, e)$ is uniformly bounded,

$$P_{\mathrm{f}}(p+n, e+n) - P_{\mathrm{f}}(p, e) \leq 2^{-p}.$$

Proof. See Section V.H.2. □

The probability of decoding failure for the initial transmission of the hybrid ARQ scheme is $P_{\mathrm{f}}(aN-K, e+(a-1)N)$, and it is $P_{\mathrm{f}}(N-K, e)$ for the standard ARQ scheme when the codeword suffers e erasures. As an immediate consequence of Proposition 102, we know that the penalty incurred in using hybrid ARQ in terms

of probability of decoding failure at the first attempt is

$$P_{\mathrm{f}}(aN-K, e+(a-1)N) - P_{\mathrm{f}}(N-K, e) \leq 2^{-(N-K)},$$

which remains very small for typical scenarios. This brings credibility to employing a punctured random code in our analysis.

Using random codes over erasure channels leads to some highly desirable properties for the hybrid ARQ problem. These properties are, in turn, instrumental in finding expressions for the probabilities of success at intermediate decoding attempts. Suppose that a codebook is generated using a $(aN-K)\times aN$ parity check matrix. For this specific code, if decoding fails given the first sN received symbols (including erasures), then it will necessarily be impossible to decode the message using the leading $(s-1)N$ received symbols. This nesting is in stark contrast to error channels.

We employ $P_{\mathrm{f}}^{(s)}(j|i)$ and $P_{\mathrm{s}}^{(s)}(j|i)$ to denote the conditional probability of decoding failure and first reliable decoding success at attempt s, respectively, with final state j and given initial state i. The conditional probabilities of decoding failure are equal to

$$P_{\mathrm{f}}^{(s)}(j|i) = \sum_{e=0}^{sN} P_{\mathrm{f}}(aN-K, e+(a-s)N) \Pr\left(E_{sN}=e, C_{sN+1}=j | C_1=i\right).$$

Above, E_{sN} represents the number of erasures over the discrete interval $[1, sN]$. Given the probabilities of failure events, the conditional probabilities of success can be evaluated in a recursive fashion. Since decoding failure and decoding success at attempt one are complementary events, we have

$$\Pr(C_{N+1}=j|C_1=i) = P_{\mathrm{f}}^{(1)}(j|i) + P_{\mathrm{s}}^{(1)}(j|i).$$

Thus, the probability of a success at time one with final state j given initial state i, can be written as

$$P_{\mathrm{s}}^{(1)}(j|i) = \Pr(C_{N+1}=j|C_1=i) - P_{\mathrm{f}}^{(1)}(j|i).$$

We note that this equation is the complement of (V.4), with a convenient new notation and appropriate parameters.

Similarly, consider the first two attempts in a hybrid ARQ scheme. Three disjoint

events can occur: decoding failure at attempt two, decoding success for the first time at attempt two, decoding success at attempt one after which the channel enters some state l. Summing over all intermediate states l,

$$\Pr(C_{2N+1}=j|C_1=i) = P_{\mathrm{f}}^{(2)}(j|i) + P_{\mathrm{s}}^{(2)}(j|i) + \sum_{l\in\mathcal{C}} P_{\mathrm{s}}^{(1)}(l|i)\Pr(C_{2N+1}=j|C_{N+1}=l).$$

Consequently, the conditional probability of being able to decode for the first time at attempt two with final state j and under initial state i is

$$P_{\mathrm{s}}^{(2)}(j|i) = \Pr(C_{2N+1}=j|C_1=i) - P_{\mathrm{f}}^{(2)}(j|i) - \sum_{l\in\mathcal{C}} P_{\mathrm{s}}^{(1)}(l|i)\Pr(C_{2N+1}=j|C_{N+1}=l).$$

Extending this procedure, we can compute the probability of a decoding success at attempt s with final state j, given initial state i,

$$P_{\mathrm{s}}^{(s)}(j|i) = \Pr\left(C_{sN+1}=j|C_1=i\right) - P_{\mathrm{f}}^{(s)}(j|i) - \sum_{r=1}^{s-1}\sum_{l\in\mathcal{C}} P_{\mathrm{s}}^{(r)}(l|i)\Pr\left(C_{sN+1}=j|C_{rN+1}=l\right).$$

This methodology provides a recursive and efficient way to compute the probabilities that, under hybrid ARQ, a system takes exactly s coded chunks to decode the original message. As in Section V.C.1, we intend to compute the matrix generating function of T, the time spent in the first level of the reduced Markov chain.

Consider the aforementioned hybrid ARQ scheme with depth equal to a. When there is a decoding failure at attempt a, the hybrid ARQ system has a few potential options. The system can discard previously received symbols altogether and start the process anew. Alternatively, the transmitter can re-encode the data segment and the information in previously received symbols can be used as side information during the decoding process. No matter what the exact strategy is, the queue occupancy of a hybrid ARQ system can always be lower and upper bounded.

- *Lower bound*: In this mode, the decoding of a message always succeeds by the ath attempt. We call this the optimistic system. Let $\check{T}$ denote the time spent in the first level of the reduced Markov chain associated with this system.

- *Upper bound*: In this mode, whenever decoding fails at the ath attempt, previously received symbols are discarded altogether and the process starts anew. We call this the pessimistic view. Let $\hat{T}$ denote the time spent in the first level

of the reduced Markov chain of this system.

In essence, $\check{T}$ and $\hat{T}$ are Markov times that provide lower and upper bounds on T, the true stopping time of the hybrid ARQ decoding process. These strategies jointly produce a near complete characterization of the behavior of hybrid ARQ systems. We turn to the specifics of the proposed approaches below.

As mentioned above, an optimistic bound (lower bound) on T can be derived using

$$\hat{P}_{\mathrm{s}}^{(a)}(j|i) = \Pr\left(C_{aN+1} = j | C_1 = i\right) - \sum_{r=1}^{a-1} \sum_{l \in \mathcal{C}} P_{\mathrm{s}}^{(r)}(l|i) \Pr\left(C_{aN+1} = j | C_{rN+1} = l\right),$$

instead of $P_{\mathrm{s}}^{(a)}(j|i)$, by assuming that the decoding always succeeds by the ath attempt. This bound holds irrespective of how the system handles failures at attempt a. We define the optimistic matrix generating function $\mathbf{G}_{\check{T}}(z) = \mathbf{G}_{\min\{T,a\}}(z)$ entrywise by

$$\left[\mathbf{G}_{\check{T}}(z)\right]_{ij} = \sum_{r=1}^{a-1} P_{\mathrm{s}}^{(r)}(j|i) z^r + \hat{P}_{\mathrm{s}}^{(a)}(j|i) z^a.$$

The pessimistic matrix generating function $\mathbf{G}_{\hat{T}}(z)$ can be derived in two steps. First, consider the matrix generating function

$$\left[\mathbf{G}_{\tilde{T}}(z)\right]_{ij} = \sum_{r=1}^{a} P_{\mathrm{s}}^{(r)}(j|i) z^r$$

Then, under the assumption that information is discarded when the a decoding attempts have failed, we get

$$\mathbf{G}_{\hat{T}}(z) = \sum_{t=0}^{\infty} z^{at} \left(\mathbf{P}_{\mathrm{f}}^{(a)}\right)^t \mathbf{G}_{\tilde{T}}(z) = \left(\mathbf{I} - z^a \mathbf{P}_{\mathrm{f}}^{(a)}\right)^{-1} \mathbf{G}_{\tilde{T}}(z).$$

Above, the matrix $\mathbf{P}_{\mathrm{f}}^{(a)}$ is defined entrywise as

$$\left[\mathbf{P}_{\mathrm{f}}^{(a)}\right]_{ij} = P_{\mathrm{f}}^{(a)}(j|i).$$

We will return to these bounds and their application in Section V.F.

V.C.3 Hitting Time to an Empty Buffer

We can build upon the matrix generating function of T to obtain the distribution of H_0. The basic insights behind this characterization are that the sojourn time at any level is finite almost surely and generating matrices can account for conditional independence.

Theorem 103: The ordinary generating function of H_0, the first-passage time to an empty queue, is given by

$$G_{H_0}(z) = \mathrm{E}\left[z^{H_q}\right] = \pi_0 \left(\mathbf{G}_T(z)\right)^m \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \tag{V.14}$$

where π_0 is the channel state probability vector at time zero.

Proof. This expression for $G_{H_0}(z)$ can be obtained from an application of mathematical induction, which proceeds backward in time. The first step consists in showing that the hypothesis holds for the base case, the sojourn time at level m,

$$\begin{aligned}
\left[\pi_0 \mathbf{G}_{T_m}(z)\right]_j &= \sum_{i=1}^{k} \left[\mathbf{G}_{T_m}(z)\right]_{ij} \Pr(C_1 = i) \\
&= \sum_{i=1}^{k} \mathrm{E}\left[z^{T_m} \mathbf{1}_{\{C_{NT_m+1}=j\}} \middle| C_1 = i\right] \Pr(C_1 = i) \\
&= \mathrm{E}\left[z^{T_m} \mathbf{1}_{\{C_{NT_m+1}=j\}}\right] = \mathrm{E}\left[z^{H_{m-1}} \mathbf{1}_{\left\{C_{NH_{m-1}+1}=j\right\}}\right]
\end{aligned}$$

where we have used the fact that $H_{m-1} = T_m$. Thus, we gather that

$$[\![z^t]\!] \left[\pi_0 \mathbf{G}_{T_m}(z)\right]_j = \Pr\left(H_{m-1} = t, C_{NH_{m-1}+1} = j\right).$$

We continue with the inductive step in a similar manner. Suppose that the hypothesis is true for a certain integer q where $0 < q < m$; that is,

$$\left[\pi_0 \mathbf{G}_{H_q}(z)\right]_j = \mathrm{E}\left[z^{H_q} \mathbf{1}_{\{C_{NH_q+1}=j\}}\right] = \left[\pi_0 \mathbf{G}_{T_m}(z) \cdots \mathbf{G}_{T_{q+1}}(z)\right]_j.$$

Then, we can write

$$\begin{aligned}
\mathrm{E}\left[z^{H_{q-1}}\mathbf{1}_{\{C_{NH_{q-1}+1}=j\}}\right] &= \mathrm{E}\left[z^{H_q+T_q}\mathbf{1}_{\{C_{NH_{q-1}+1}=j\}}\right] \\
&= \sum_{i=1}^{k} \mathrm{E}\left[z^{H_q+T_q}\mathbf{1}_{\{C_{NH_{q-1}+1}=j\}}\middle| C_{NH_q+1}=i\right] \Pr(C_{NH_q+1}=i) \\
&= \sum_{i=1}^{k} \mathrm{E}\left[z^{H_q}\mathbf{1}_{\{C_{NH_q+1}=i\}}\right] \mathrm{E}\left[z^{T_q}\mathbf{1}_{\{C_{NH_{q-1}+1}=j\}}\middle| C_{NH_q+1}=i\right] \\
&= \sum_{i=1}^{k} \left[\pi_0 \mathbf{G}_{T_m}(z)\cdots\mathbf{G}_{T_{q+1}}(z)\right]_i \left[\mathbf{G}_{T_q}(z)\right]_{ij} \\
&= \left[\pi_0 \mathbf{G}_{T_m}(z)\cdots\mathbf{G}_{T_q}(z)\right]_j = \left[\pi_0 \mathbf{G}_{H_{q-1}}(z)\right]_j .
\end{aligned}$$

That is, the hypothesis is also true for $q-1$. We note that the third equality follows from the conditional independence of our quasi-birth-death Markov process. In our problem, we have $\mathbf{G}_{T_q}(z) = \mathbf{G}_T(z)$ for all $q \in \{1,\ldots,m\}$. Since this expression holds for any π_0, we conclude that $\mathbf{G}_{H_0}(z) = (\mathbf{G}_T(z))^m$ and, as a consequence,

$$[\![z^t]\!]\left[\pi_0 \left(\mathbf{G}_T(z)\right)^m\right]_j = \Pr\left(H_0 = t, C_{NH_0+1} = j\right).$$

Summing over all the possible end states, we recover the expression for $G_{H_0}(z)$ given in (V.14). □

To differentiate among possible initial conditions, it will become useful to write the first-passage time to an empty queue with an initial buffer size of m segments as $H_0^{(m)}$.

V.D LARGE DEVIATION ANALYSIS

As seen in the previous section, it is possible to evaluate the exact distribution of $H_0^{(m)}$. This facilitates the selection of parameters to optimize overall performance. However, this process becomes cumbersome for large buffer sizes. In such circumstances, analyzing the large deviations governing the system offers a new direction to derive meaningful guidelines for resource allocation and parameter tuning. Below, we study two types of aberrations under the ARQ scheme: deviations in the average transmission time and the mean service rate. We note that, although large deviations can be studied under hybrid ARQ, this latter scenario is somewhat tedious and it offers limited additional insights. Hence we restrict our attention to the ARQ scheme.

We begin with the average transmission time; that is, the normalized first-passage time to an empty queue.

V.D.1 Normalized First-Passage Time

Again, suppose that the transmit buffer contains exactly m segments at the onset of the communication process. We are interested in the large deviations associated with the sequence of random variables specified by

$$Y_m = \frac{1}{m} H_0^{(m)} = \frac{1}{m} \sum_{q=1}^{m} T_q \quad m = 1, 2, \ldots$$

The logarithmic moment generating function for Y_m is

$$\begin{aligned} \Lambda_m(\lambda) &= \log \mathrm{E}\left[e^{\lambda Y_m}\right] = \log \mathrm{E}\left[e^{\lambda H_0^{(m)}/m}\right] \\ &= \log G_{H_0}^{(m)}\left(e^{\lambda/m}\right). \end{aligned}$$

The existence of limits of properly scaled logarithmic moment generating functions suggests that $\{Y_m\}$ may satisfy a large deviation principle [182]. In particular, consider the following asymptotic regime

$$\begin{aligned} \Lambda(\lambda) &= \lim_{m\to\infty} \frac{1}{m} \Lambda_m(m\lambda) = \lim_{m\to\infty} \frac{1}{m} \log G_{H_0}^{(m)}\left(e^{\lambda}\right) \\ &= \lim_{m\to\infty} \frac{1}{m} \log\left(\pi_0 \left(\mathbf{G}_T\left(e^{\lambda}\right)\right)^m \mathbf{1}\right). \end{aligned} \tag{V.15}$$

A few observations concerning $\Lambda(\lambda)$ are in order. In view of Lemma 101 and for $z = e^{\lambda}$,

$$\mathbf{G}_T\left(e^{\lambda}\right) = \left(\sum_{t=0}^{\infty} \mathbf{K}^t e^{t\lambda}\right) \mathbf{M} e^{\lambda}$$

is a non-negative matrix over the extended real numbers. In fact, this matrix possesses additional properties which are summarized below. Again, let $\varrho(\cdot)$ denote the spectral radius of its matrix argument.

Lemma 104: If T is finite almost surely, the matrix generator $\mathbf{G}_T\left(e^{\lambda}\right)$ exists as a non-negative real matrix if and only if $\lambda < -\log \varrho(\mathbf{K})$. In particular, when $\lambda \geq -\log \varrho(\mathbf{K})$, one or more entries of $\mathbf{G}_T(e^{\lambda})$ will be infinite.

Proof. See Section V.H.3. □

Another important quantity is the spectral radius of $\mathbf{K}$, which is related to the support of $\mathbf{G}_T\left(e^{\lambda}\right)$ as seen in Lemma 104.

Corollary 105: If T is finite almost surely, then $\varrho(\mathbf{K}) < 1$.

Proof. See Section V.H.4. □

Under Assumption 98 and for any non-trivial coding scheme, T is finite almost surely, thus the hypotheses of Lemma 104 and Corollary 105 are satisfied. A sufficient condition to ensure the existence of a large deviation principle for the average transmission time is that the Markov process $\{U_t\}$ sampled at departure events $\{H_q\}$ is irreducible. This guarantees that the states of the corresponding jump chain form a unique recurrent class. Formally, we postulate the following condition.

Assumption 106: The matrix $(\mathbf{I}-\mathbf{K})^{-1}\mathbf{M}$ is irreducible.

We note that, strictly speaking, this is not a necessary condition. Having a unique communicating class and, possibly, transient states in the jump chain will also work. However, this more encompassing setting leads to extra bookkeeping, which unnecessarily clouds some of the underlying concepts. Furthermore, all the practical systems we wish to study fulfill the requirements of Assumption 106. As such, we take it for granted from this point forward. Under this assumption, the matrix $\mathbf{G}_T\left(e^{\lambda}\right)$ is irreducible for any $\lambda < -\log\varrho(\mathbf{K})$ and, hence, the Perron-Frobenius theorem applies [182, 196]. This leads to the following result.

Proposition 107: Under Assumption 106, the limiting moment generating function defined in (V.15) exists as an extended real number for every $\lambda \in \mathbb{R}$, with

$$\Lambda(\lambda) = \begin{cases} \varrho\left(\left(\mathbf{I}-\mathbf{K}e^{\lambda}\right)^{-1}\mathbf{M}e^{\lambda}\right) & \lambda < -\log\varrho(\mathbf{K}) \\ \infty & \text{otherwise.} \end{cases}$$

Proof. See Section V.H.5. □

Using matrix norms, it can be shown that $\mathbf{G}_T\left(e^{\lambda}\right)$ is differentiable entrywise over the interval $\lambda < -\log\varrho(\mathbf{K})$. Since $\Lambda(\lambda)$ is an isolated root of the characteristic function of matrix $\mathbf{G}_T\left(e^{\lambda}\right)$, we deduce that it is positive, finite and differentiable with respect to λ (see, e.g., [197, Th. 11.5.1], [182, p. 75]). Corollary 105 asserts that $\varrho(\mathbf{K}) < 1$, which implies that $\Lambda(0)$ is finite. In view of the discussion above, we

conclude that the origin is in the interior of $\{\lambda \in \mathbb{R} : \Lambda(\lambda) < \infty\}$. Consequently, $\Lambda(\lambda)$ is essentially smooth and the Gärtner-Ellis theorem applies [182], thereby establishing the desired result.

Theorem 108: Suppose $\left\{Y_m = \frac{1}{m}\sum_{q=1}^{m} T_q\right\}$ is the empirical mean sojourn time per level. For every $x \in \mathbb{R}$, consider the Fenchel-Legendre transform

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \left\{\lambda x - \log \varrho\left(\mathbf{G}_T\left(e^{\lambda}\right)\right)\right\}. \tag{V.16}$$

The empirical mean Y_m satisfies the large deviation principle with the convex, good rate function $\Lambda^*(\cdot)$. That is, for any set $\Gamma \subseteq \mathbb{R}$ and any initial state $c \in \mathcal{C}$,

$$\begin{aligned} -\inf_{x \in \Gamma^o} \Lambda^*(x) &\leq \liminf_{m\to\infty} \frac{1}{m} \log \Pr(Y_m \in \Gamma) \\ &\leq \limsup_{m\to\infty} \frac{1}{m} \log \Pr(Y_m \in \Gamma) \leq -\inf_{x \in \bar{\Gamma}} \Lambda^*(x), \end{aligned}$$

where Γ^o and $\bar{\Gamma}$ denote the interior and closure of the set Γ, respectively.

Example 109: For the Gilbert-Elliott channel shown in Fig. V.1, it is possible to obtain a closed-form expression for the spectral radius of $\mathbf{G}_T\left(e^{\lambda}\right)$. Specifically, we can write the characteristic polynomial of $\mathbf{G}_T\left(e^{\lambda}\right)$ as

$$\begin{aligned} \det\left(\gamma\mathbf{I} - \mathbf{G}_T\left(e^{\lambda}\right)\right) &= \det\left(\gamma\mathbf{I} - \left(\mathbf{I} - \mathbf{K}e^{\lambda}\right)^{-1}\mathbf{M}e^{\lambda}\right) \\ &= \frac{\det\left(\gamma\mathbf{I} - \gamma\mathbf{K}e^{\lambda} - \mathbf{M}e^{\lambda}\right)}{\det\left(\mathbf{I} - \mathbf{K}e^{\lambda}\right)}. \end{aligned}$$

We note that the numerator is a quadratic equation in γ and the denominator is a constant. It is therefore possible to find parametric expressions for the two roots of $\det\left(\gamma\mathbf{I} - \mathbf{G}_T\left(e^{\lambda}\right)\right)$. Taking the maximum of the absolute values of these two roots yields an explicit, albeit convoluted, expression for the spectral radius of $\mathbf{G}_T\left(e^{\lambda}\right)$. As such, $\Lambda^*(\cdot)$ can be obtained efficiently.

V.D.2 Empirical Mean Service

We turn to the second type of aberrations we wish to study: deviations in the empirical mean service rate,

$$Z_s = \frac{1}{s}\sum_{t=1}^{s} D_t.$$

We note that $\{D_s\}$ is not a Markov process. However, D_s is a (trivial) deterministic function of $V_s = (C_{(s+1)N+1}, D_s)$. Since $\{V_s\}$ is a Markov process, we can apply general results on the large deviation principle of additive functionals of Markov chains. To leverage these results, we first impose an ordering on the state space $\mathcal{V} = \mathcal{C} \times \{0, 1\}$. Recall that $|\mathcal{C}| = k$; a natural ordering for this state space is to associate integer $v = (dk + i)$ with state (i, d). Using this ordering, the transition probability matrix $\mathbf{\Pi}$ for the augmented process $\{V_s\}$ is given by

$$[\mathbf{\Pi}]_{v_1,v_2} = \pi(v_1, v_2), \quad v_1, v_2 \in \{1, \ldots, 2k\}$$

where $\pi(v_1, v_2)$ is the probability of jumping to state v_2, conditioned on starting from v_1.

Assumption 110: The matrix $\mathbf{\Pi}$ is irreducible.

This assumption is similar in spirit to Assumption 106. Yet the large deviation principle on the empirical service can be derived under weaker conditions. In particular, it suffices to show that $\mathbf{K} + \mathbf{M}$ is irreducible, a requirement that is easily met. We stress that $\mathbf{K} + \mathbf{M}$ is equal to $\mathbf{B}^N$, and the latter matrix is itself irreducible by Assumption 98.

Theorem 111 ([182])**:** Let $\{V_s\}$ be a finite-state Markov chain possessing an irreducible transition matrix $\mathbf{\Pi}$. For every $x \in \mathbb{R}$, define

$$I(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \log \varrho\,(\mathbf{\Pi}_\lambda)\} \tag{V.17}$$

where $\mathbf{\Pi}_\lambda$ is a nonnegative matrix whose elements are

$$\pi_\lambda\,(v_1, v_2) = \pi(v_1, v_2) e^{\lambda d_2} \quad v_1, v_2 \in \{1, \ldots, 2k\}.$$

Then, the empirical mean Z_s satisfies the large deviation principle with the convex good rate function $I(\cdot)$. Explicitly, for any set $\Gamma \subseteq \mathbb{R}$, and any initial state $v \in \mathcal{V}$,

$$\begin{aligned} -\inf_{x \in \Gamma^o} I(x) &\leq \liminf_{s \to \infty} \frac{1}{s} \log P_v^\pi(Z_s \in \Gamma) \\ &\leq \limsup_{s \to \infty} \frac{1}{s} \log P_v^\pi(Z_s \in \Gamma) \leq -\inf_{x \in \bar{\Gamma}} I(x) \end{aligned}$$

where P_v^π denotes the Markov probability measure induced by transition probability $\mathbf{\Pi}$ and initial state $v \in \mathcal{V}$, i.e.,

$$P_v^\pi(V_1 = v_1, \ldots, V_s = v_s) = \pi(v, v_1) \prod_{t=1}^{s-1} \pi(v_t, v_{t+1}).$$

Expressions for the transition probabilities used in this theorem appear in (V.6). We note that

$$\Pr\left(V_{s+1} = (j, d_2) | V_s = (i, d_1)\right) = \Pr\left(V_{s+1} = (j, d_2) | C_{(s+1)N+1} = i\right);$$

this induces a repetitive structure in matrix $\mathbf{\Pi}$. The nonnegative matrix $\mathbf{\Pi}_\lambda$ associated with every $\lambda \in \mathbb{R}$ can then be written explicitly as

$$\mathbf{\Pi}_\lambda = \begin{bmatrix} \kappa_{11} & \cdots & \kappa_{1k} & \mu_{11}e^\lambda & \cdots & \mu_{1k}e^\lambda \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \kappa_{k1} & \cdots & \kappa_{kk} & \mu_{k1}e^\lambda & \cdots & \mu_{kk}e^\lambda \\ \kappa_{11} & \cdots & \kappa_{1k} & \mu_{11}e^\lambda & \cdots & \mu_{1k}e^\lambda \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \kappa_{k1} & \cdots & \kappa_{kk} & \mu_{k1}e^\lambda & \cdots & \mu_{kk}e^\lambda \end{bmatrix}. \tag{V.18}$$

We can rewrite $\mathbf{\Pi}_\lambda$ by taking advantage of its block structure,

$$\mathbf{\Pi}_\lambda = \begin{bmatrix} \mathbf{K} & \mathbf{M}e^\lambda \\ \mathbf{K} & \mathbf{M}e^\lambda \end{bmatrix}.$$

The pertinent eigenvalues are the roots of the characteristic polynomial of $\mathbf{\Pi}_\lambda$. Using properties of matrix determinant and the commutative properties of some of the blocks, we can express this polynomial as

$$\begin{aligned} \det\left(\gamma\mathbf{I} - \mathbf{\Pi}_\lambda\right) &= \det\left((\gamma\mathbf{I} - \mathbf{M}e^\lambda)(\gamma\mathbf{I} - \mathbf{K}) - \mathbf{M}\mathbf{K}e^\lambda\right) \\ &= \det\left((\gamma\mathbf{I} - \mathbf{K})(\gamma\mathbf{I} - \mathbf{M}e^\lambda) - \mathbf{K}\mathbf{M}e^\lambda\right) \\ &= \det\left(\gamma^2\mathbf{I} - \gamma\mathbf{K} - \gamma\mathbf{M}e^\lambda\right). \end{aligned}$$

Collectively, Theorem 111 and the matrix defined in (V.18) provide an algorithmic workflow for the computation of the good rate function associated with the empirical

means $\{Z_s\}$. We follow this discussion with an example based on a two-state channel with memory.

Example 112: Once again, consider a Gilbert-Elliott erasure channel with $\mathcal{C} = \{1, 2\}$. An advantage in studying this rudimentary model is that it admits a simple, closed-form characterization. The dimension of the state space in this case is $|\mathcal{V}| = 4$. Using the commutative block structure discussed above, the determinant of $(\gamma\mathbf{I} - \mathbf{\Pi}_\lambda)$ reduces to

$$\begin{aligned}\det\left(\gamma\mathbf{I} - \mathbf{\Pi}_\lambda\right) &= \det\left(\gamma^2\mathbf{I} - \gamma\mathbf{K} - \gamma\mathbf{M}e^\lambda\right) \\ &= \gamma^2 \det\left(\begin{bmatrix} \gamma - \kappa_{11} - \mu_{11}e^\lambda & -\kappa_{12} - \mu_{12}e^\lambda \\ -\kappa_{21} - \mu_{21}e^\lambda & \gamma - \kappa_{22} - \mu_{22}e^\lambda \end{bmatrix}\right).\end{aligned}$$

By inspection, we see that the spectral radius of $\mathbf{\Pi}_\lambda$ is the largest root of the quadratic equation

$$\begin{aligned}&\gamma^2 - \gamma(\kappa_{11} + \kappa_{22} + (\mu_{11} + \mu_{22})e^\lambda) \\ &\quad + (\kappa_{11} + \mu_{11}e^\lambda)(\kappa_{22} + \mu_{22}e^\lambda) - (\kappa_{12} + \mu_{12}e^\lambda)(\kappa_{21} + \mu_{21}e^\lambda) = 0.\end{aligned}$$

For fixed parameters, this dominating root can be computed using the celebrated quadratic formula. We will revisit this example in Section V.F.

V.D.3 Relation between $\Lambda^*(\cdot)$ and $I(\cdot)$

The two rate functions introduced above, $\Lambda^*(\cdot)$ and $I(\cdot)$, characterize the large deviation principles for the mean transmission time and average service rate, respectively. Since the processes $\{T_q\}$ and $\{D_s\}$ are closely related, one can presume that their governing rate functions are somehow linked. A key insight in understanding this relation is to realize that the following events are equivalent: for any positive integers m and n,

$$\{T_1 + \cdots + T_m > n\} = \{D_1 + \cdots + D_n < m\}. \tag{V.19}$$

In words, the first event occurs whenever more than n attempts are required to successfully deliver m packets, while the second event states that fewer than m packet transmissions have been successful within the first n attempts. Using this relationship and scaling arguments, one can establish our next proposition which

substantiates the existence of a strong connection between the two rate functions.

Proposition 113: If the rate functions $\Lambda^*(\cdot)$ and $I(\cdot)$ are finite in the open intervals $(1, \infty)$ and $(0, 1)$, respectively, then they satisfy

$$I(x) = x\Lambda^* \left(\frac{1}{x} \right)$$

for $x \in (0, 1)$.

Proof. See Section V.H.6. □

V.E PERFORMANCE EVALUATION

Thus far, we have devoted much attention to developing a thorough understanding of H_0 and, in particular, its generating function. In this section, we apply the results of Theorem 103 and we derive a number of pertinent performance criteria with practical significance.

First, recall that $[\![z^t]\!] G_{H_0}(z) = \Pr(H_0 = t)$. Accordingly, the probability that the queue fails to drain within τ time units is equal to

$$\Pr(H_0 > \tau) = 1 - \sum_{t=0}^{\lfloor \tau \rfloor} [\![z^t]\!] G_{H_0}(z).$$

Moreover, the average time required to empty the queue is obtained by differentiating the moment generating function of H_0 and then taking the limit as z approaches one,

$$\mathrm{E}\left[H_0\right] = \lim_{z \uparrow 1} \frac{d}{dz} G_{H_0}(z).$$

Alternatively, using Chernoff inequalities, it is possible to upper bound the probability of a deviation event in a computationally efficient manner. The equation

$$\Pr(H_0 > \tau) \leq e^{-\lambda\tau} \mathrm{E}\left[e^{\lambda H_0}\right] = e^{-\lambda\tau} G_{H_0}\left(e^{\lambda}\right)$$

holds for any $\lambda > 0$. The optimal bound derived from this collection of inequalities is sometimes expressed in logarithmic form,

$$\log \Pr(H_0 > \tau) \leq - \sup_{\lambda > 0} \left\{ \lambda\tau - \log\left(G_{H_0}\left(e^{\lambda}\right) \right) \right\}.$$

The large deviation principle on H_0 derived in Section V.D confirms that, under mild conditions, this latter bound is asymptotically tight.

It may be instructive to stress that H_0, the first-passage time introduced in (V.7), is defined in terms of codeword transmission attempts. That is, H_0 represents the cumulative number of codewords sent by the source until the queue empties out completely. Such a metric poses no issue when comparing systems of identical block lengths. However, when assessing the performance of candidate implementations with different block lengths, a more careful interpretation of the results becomes necessary. This subtlety arises because of the mismatch in indexing between the evolution of the queue and the number of channel uses. For a fair evaluation of potential candidates, hitting times should be scaled to portray their evolution according to a common clock, that of the channel process.

Define random variable $\tilde{H}_0$ by

$$\tilde{H}_0 = NH_0,$$

where N designates the block length associated with the underlying implementation. Then, $\tilde{H}_0$ denotes the number of channel uses necessary to empty out the queue, and it can therefore be employed to provide a uniform measure of performance. While it is straightforward to extend our performance criteria to $\tilde{H}_0$ through the relation

$$\Pr(H_0 > \tau) = \Pr\left(\tilde{H}_0 > \frac{\tau}{N}\right),$$

it is essential to apply this transformation when comparing systems with different block lengths.

A similar scaling is needed when comparing the large deviations of systems with different parameters. A proper scaling for the fair comparison of mean sojourn times can be expressed in terms of channel uses per information bit,

$$\tilde{Y}_\ell = \frac{1}{\ell} N H_0^{(\lceil \ell/K \rceil)}.$$

This leads to the following asymptotic regime

$$\lim_{\ell\to\infty} \frac{1}{\ell} \log \Pr\left(\tilde{Y}_\ell > \tau\right) = \frac{1}{K} \lim_{\ell\to\infty} \frac{1}{\lceil \ell/K \rceil} \log \Pr\left(\frac{1}{\lceil \ell/K \rceil} H_0^{(\lceil \ell/K \rceil)} > \frac{K}{N}\tau\right)$$

$$= \frac{1}{K} \lim_{m\to\infty} \frac{1}{m} \log \Pr\left(\frac{1}{m} H_0^{(m)} > \frac{K}{N}\tau\right)$$
$$= -\frac{1}{K}\Lambda^*\left(\frac{K}{N}\tau\right)$$

where $\tau > \mathrm{E}\left[\tilde{Y}_\infty\right]$. Likewise, to account for discrepancies in design parameters, the empirical mean service can be expressed in terms of decoded bits per channel use,

$$\tilde{Z}_n = \frac{1}{n} \sum_{t=1}^{\lfloor n/N \rfloor} K D_t.$$

The ensuing asymptotic regime becomes

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr\left(\tilde{Z}_n < \eta\right) = \frac{1}{N} \lim_{n\to\infty} \frac{1}{\lfloor n/N \rfloor} \log \Pr\left(\frac{1}{\lfloor n/N \rfloor} \sum_{t=1}^{\lfloor n/N \rfloor} D_t < \frac{N}{K}\eta\right)$$
$$= \frac{1}{N} \lim_{s\to\infty} \frac{1}{s} \log \Pr\left(\frac{1}{s} \sum_{t=1}^{s} D_t < \frac{N}{K}\eta\right)$$
$$= -\frac{1}{N} I\left(\frac{N}{K}\eta\right)$$

where $\eta < \mathrm{E}\left[\tilde{Z}_\infty\right]$. Collectively, these various modifications enables the comparison of competing implementations with different values for K and N.

Another concern that comes into play when optimizing over block length is the impact of the initial state of the system. If the number of bits at the source is fixed at time zero, the scope of the optimal solution may be very narrow. This is a situation akin to over-fitting in statistical modeling. To provide a more robust characterization with widely applicable results and guidelines, it may be beneficial to assume that the number of bits in the queue at the onset of the transmission process is random, with a prescribed representative distribution. In our numerical study, we circumvent some of these difficulties by assuming that the block length is fixed and the initial queue length is random. The specifics of our investigation are detailed below.

V.F NUMERICAL ANALYSIS

In this section, we apply the methodology developed above to an illustrative example. Physical parameters are selected to resemble an implementation of the

global system for mobile communications (GSM). Specifically, the block length is fixed at $N = 114$. The information content per codeword, K, is a parameter to be optimized. We model the wireless connection as a Gilbert-Elliott erasure channel, and we denote its transition probability matrix as

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

For simplicity, we assume that $\varepsilon_1 = 1$ and $\varepsilon_2 = 0$. The probability of a bit erasure is set at twenty percent, which entails

$$\frac{b_{21}}{b_{12} + b_{21}} = 0.2.$$

For this elementary model, channel memory can be expressed unambiguously through the decay factor $(1 - b_{12} - b_{21})$, which is determined by the spectrum of the matrix. A decay factor equal to zero is equivalent to a memoryless channel, while correlation increases as $(1 - b_{12} - b_{21})$ approaches one. Except where specified otherwise, we employ a decay factor equal to 0.9 in our numerical results.

We assume that L, the number of information bits contained at the source at time zero, is a random variable possessing a Gamma distribution with mean 2000 and standard deviation 100. Randomizing the number of bits at the source partly alleviates the idiosyncratic effects associated with partitioning the queue content into segments of K bits. For a source buffer with ℓ information bits, the number of segments to be delivered is $\lceil \ell/K \rceil$ and, as such, a one-bit variation in ℓ can result in having an additional message to send. Imposing a random distribution on the number of information bits at the source leads to a probability distribution on $M = \lceil L/K \rceil$. This, in turn, yields smoother results.

Figures V.4 and V.5 present the mean and variance of the first-passage times for the ARQ and hybrid ARQ schemes as functions of the number of information bits per codeword. Varying the code rate affects both the expected value of the first-passage time and its variance. A low code rate offers more protection against erasures and, accordingly, the resulting distribution of the hitting time to an empty queue is very narrow. Increasing the code rate initially reduces the mean first-passage time, as every successful decoding attempt reveals more information bits. However, a higher code rate also raises the probability of decoding failure. Eventually, as the code rate

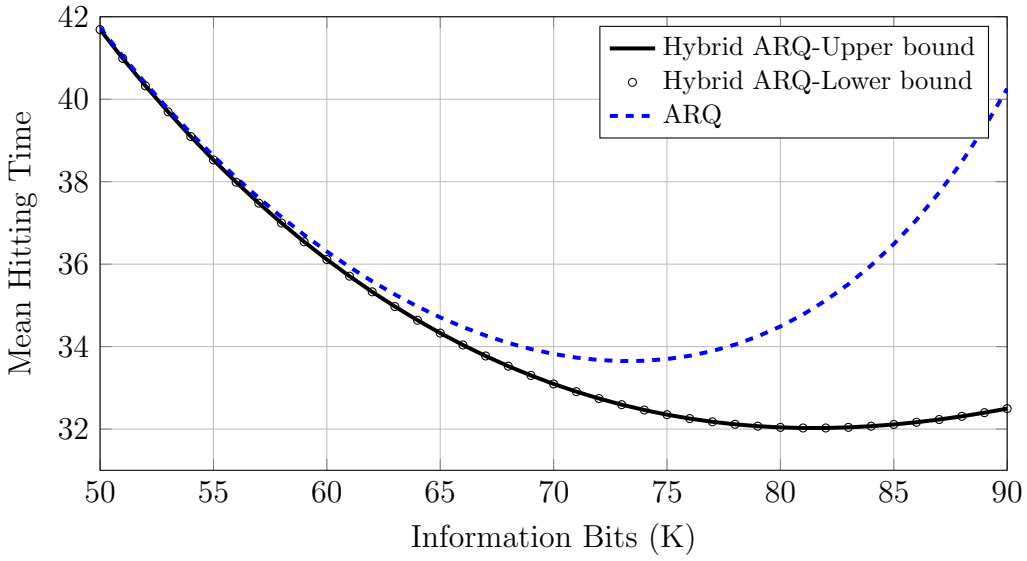

Figure V.4: This figure shows mean first-passage times as functions of K. The block length employed in all cases is $N = 114$. The underlying Gilbert-Elliott channel produces erasures with probability 0.20, and it possesses a dominant decay factor of $(1 - b_{12} - b_{21}) = 0.9$. The expected number of bits at the source at time zero is 2000. The upper and lower bounds for the hybrid ARQ scheme with a depth of $a = 3$ are indistinguishable.

is pushed further, decoding failures start to hamper the draining process and the mean first-passage time grows due to excessive repetition requests. This effect is much more pronounced for standard ARQ.

The penalty in using a high code rate is less severe for the hybrid ARQ scheme because the failure recovery mechanism, which is based on incremental redundancy, adapts gracefully to channel conditions in this latter case. For instance, when K is very close to N, decoding under standard ARQ will fail nearly every time. Contrastingly, the effective code rate drops rapidly with decoding failures under hybrid ARQ. The robust profile of hybrid ARQ is a key property that underlies the popularity of this paradigm in practical systems. In the current example, the upper and lower bounds derived for $\mathrm{E}[H_0]$ under the hybrid ARQ scheme are essentially indistinguishable, hinting at the fact that decoding failures are nearly nonexistent once three blocks are received.

Perhaps not too surprisingly, our numerical investigation suggests that the optimal code rate is somewhat impervious to initial queue conditions. To examine

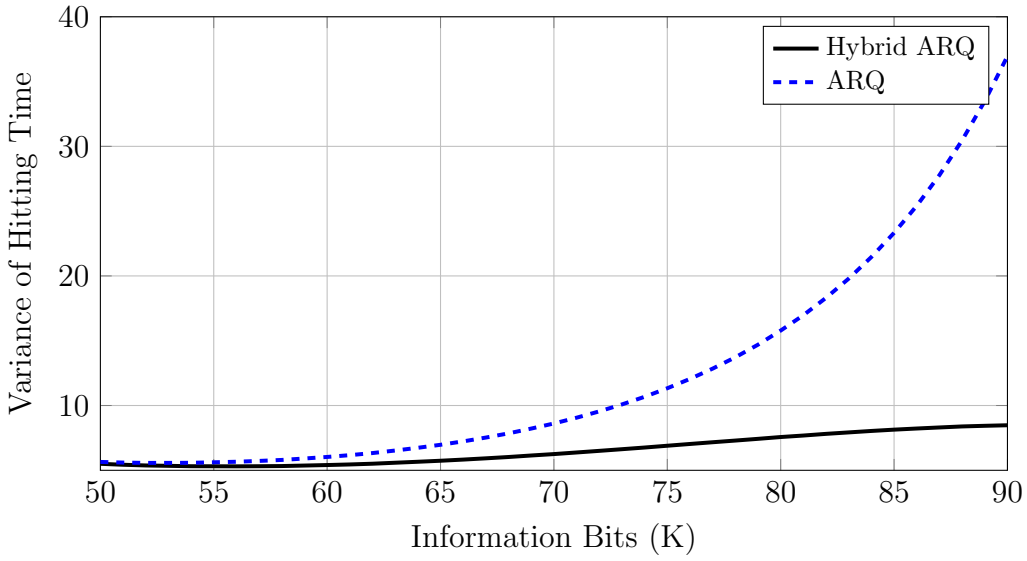

Figure V.5: This figure displays variances of the first-passage times to an empty queue as functions of K. The parameters used in this numerical study are the same as those featured in Fig. V.4. The variance for the hybrid ARQ scheme is calculated with the upper bound $\hat{T}$.

---

the effects of the initial queue length, we employ the channel parameters described above and we modify the distribution on L. For Gamma distributions with means $\mathrm{E}[L] \in \{500, 1000, 2000, 3000\}$ and standard deviation 100, the optimal value of K in terms of mean first-passage time is consistently equal to 73 for standard ARQ and it remains fixed at 81 for the hybrid variant.

Using the methodology established thus far, it is possible to consider additional performance criteria. For instance, we can analyze the crossings of the cumulative distribution function,

$$h_p = \min_t \{t | \Pr(H_0 \leq t) \geq p\}.$$

Fig. V.6 plots the number of transmission attempts associated with threshold values $p \in \{0.45, 0.95\}$. We observe that the optimal value of K decreases slightly when the crossing threshold p approaches one. In other words, when focusing on worst-case behavior, the system tends to favor a more conservative setting with extra protection against erasures. This phenomenon offers another perspective on the tradeoff between expected behavior and its variations.

Next, we turn to the large deviations techniques developed in Section V.D. As

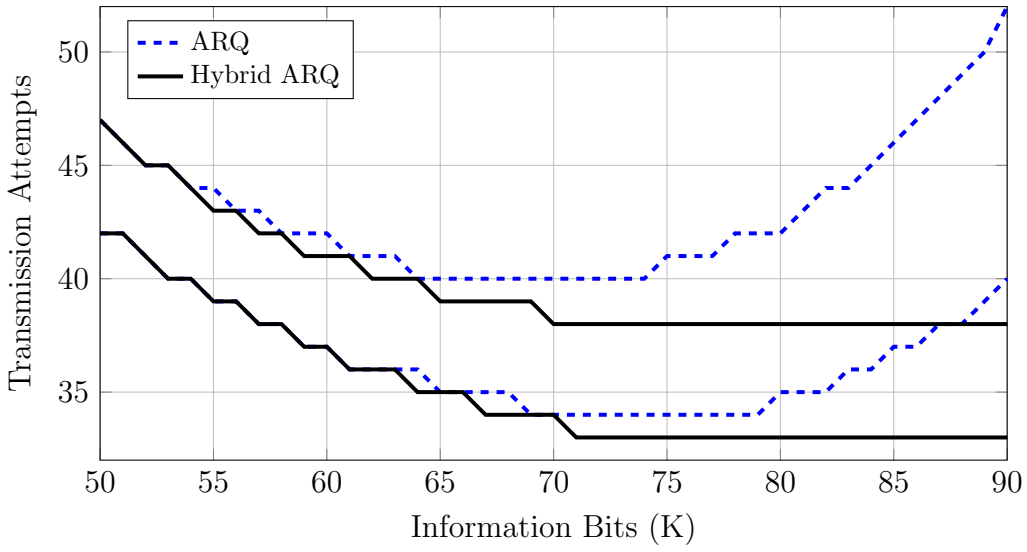

Figure V.6: The crossings of the cumulative distribution function $F_{H_0}(\cdot)$ offer conservative figures of merit for the operation of the queueing system. In this example, the lines correspond to thresholds $p \in \{0.45, 0.95\}$.

---

a reference, we consider a voice stream application. In GSM, each speech frame of length 20 ms is encoded into a data segment of length 228. The underlying physical layer has the ability to transmit one symbol every 40 μs. If we approximate the maximum delay tolerance for one-way voice traffic to be 40 ms [198, p. 70], then this requires 228 bits to be transmitted within roughly 1000 channel uses. This constraint, in turn, necessitates a nominal rate on the order of 0.23 bits per channel use for link reliability. We adopt this figure as a rough estimate for the needs of a voice stream in our numerical study.

The maximum throughput that can be supported over the Gilbert-Elliott channel in our example is slightly above 0.5 bits per channel use. Recall that threshold η represents a minimum target requirement on the number of information bits per channel use that can be successfully decoded at the destination, in an asymptotic regime. When $\eta < 0.5$, there exist values of K for which the rate function $\frac{1}{N} I\left(\frac{N}{K}\eta\right)$ is strictly positive; this can be seen in Fig. V.7. These curves can be used to characterize the tension between quantization and failures to deliver media properly. A high-quality stream, with a large η, will offer an enhanced viewer experience when transmitted adequately, but will necessarily be more prone to interruptions and failures, as exposed through the rate functions. A low-bandwidth, low-quality stream

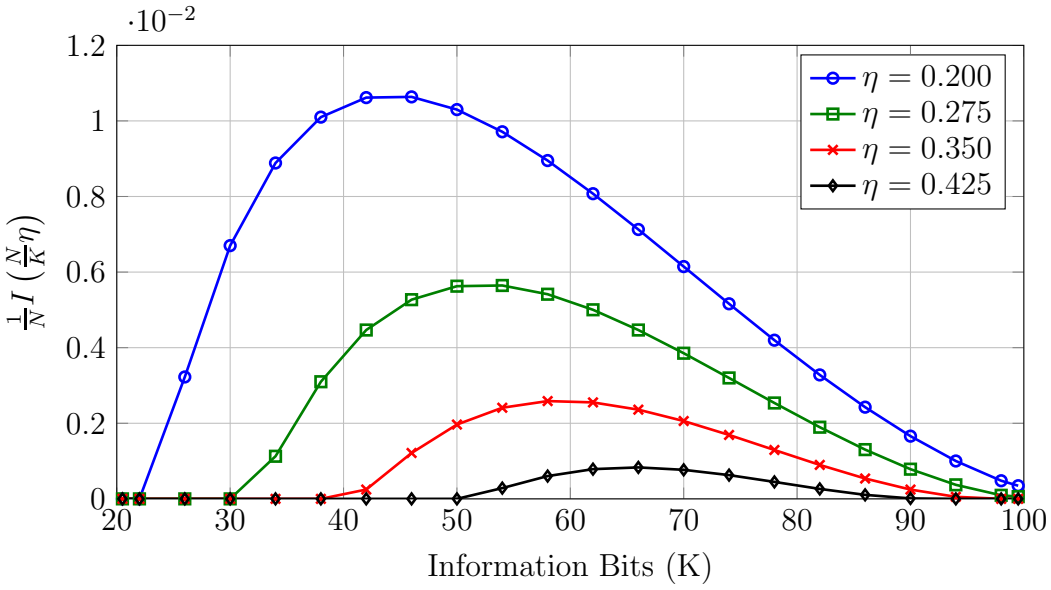

Figure V.7: This figure plots good rate functions governing large deviations in the empirical mean service as functions of K, the number of information bits per codeword. Given throughput threshold η, the optimal value of K is the argument corresponding to the apex of the function.

---

on the other hand offers a better delivery profile with a smaller probability of failure. However, the quality of the playback may not be satisfactory to the end user. A proper selection of parameters for an adequate overall user experience can be made through the rate functions of Fig. V.7.

Once η is picked, the corresponding curve displays performance as a function of K. For low code rates, the maximum achievable throughput is less than the service requirement and hence the rate function governing large deviations is zero. At high code rates, performance is limited by the rise in the probability of decoding failure. The system must then find the right balance between the frequency of failures and the payoff of a decoding success in terms of information bits. The optimal value of K for a specific threshold η is given by the apex of its curve,

$$K_Z^*(\eta) = \arg\max_K \frac{1}{N} I\left(\frac{N}{K}\eta\right).$$

It is interesting to note how conservative the optimal code rate becomes when the target service requirement is reduced.

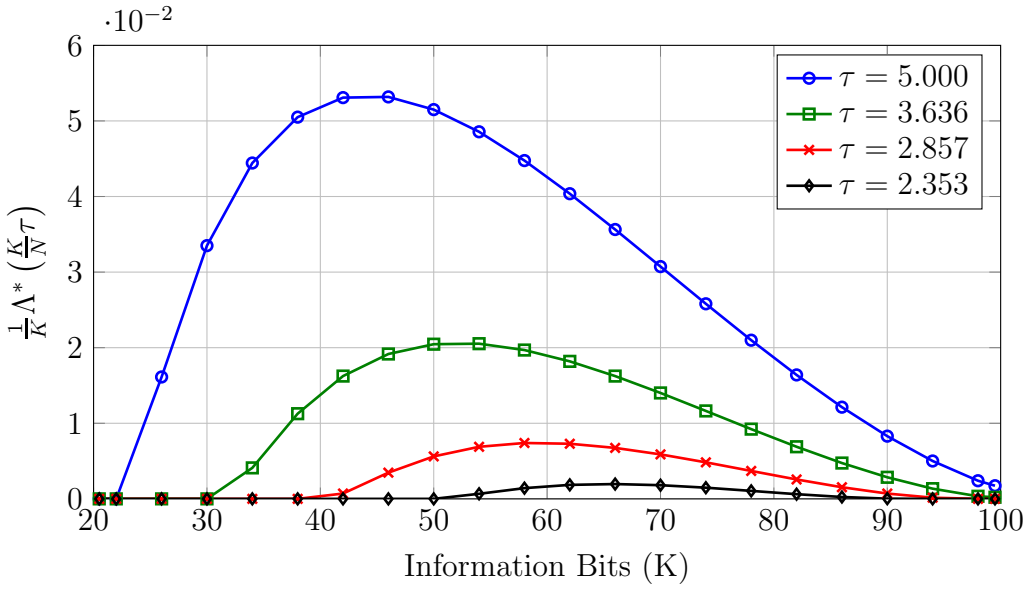

Figure V.8: This figure shows good rate functions governing large deviations in the mean sojourn time as functions of K. The optimum code rate depends heavily on the deviation threshold of the mean sojourn time.

The second type of rate functions introduced in Section V.D characterizes large deviations in the mean sojourn times, as shown in Fig. V.8. These curves can be employed to tradeoff playback quality and buffering times for streaming media. More specifically, τ represents a limitation on the average number of channel uses employed to transmit one bit of information. Of course, when a high-quality rendering is selected, the system must deliver a larger amount of data within the buffering window and, hence, the probability of delay violation becomes greater. In this case, the optimal value of K becomes

$$K_Y^*(\tau) = \arg\max_K \frac{1}{K}\Lambda^*\left(\frac{K}{N}\tau\right).$$

The behavior of the system in terms of average sojourn time is closely related to the empirical mean service, holding a reciprocal relation. We emphasize that the optimal code rates are equal, namely $K_Z^*(\eta) = K_Y^*(\tau)$ whenever $\tau = \eta^{-1}$. This is due to the relation between $I(\cdot)$ and $\Lambda^*(\cdot)$ described in Section V.D.3.

The last aspect of this system we wish to explore is the potential impact of channel memory and correlation among successive channel uses. As before, we keep the

Channel Memory	Optimal Value of K		Mean First-Passage Time E$[H_0]$		Crossing $h_{0.95}$	
	ARQ	HARQ	ARQ	HARQ	ARQ	HARQ
0	81	81	26.92	25.90	30	30
0.5	77	78	28.87	27.79	32	32
0.9	73	81	34.65	32.03	40	38
0.95	77	96	36.68	31.75	44	38
0.98	95	107	35.21	28.62	45	36

Table V.1: Optimal number of information bits per codeword as a function of channel memory factor $1 - b_{12} - b_{21}$.

probability of a bit erasure at twenty percent. However, we vary the decay factor of the channel, $(1-b_{12}-b_{21})$, from zero to one. Once again, we assess performance using the mean first-passage time to an empty queue. When the channel is memoryless, the optimal value for K is 81. As correlation increases, more protection against erasures is beneficial and the optimal value of K decreases moderately. This enables the system to compensate for short sequences of erasures. Still, as correlation strengthens, it becomes difficult to correct longer strings of erasures. When this happens, the penalty of a smaller payoff produced by a low rate code begins to dominate. In other words, attempting to recover every packet starts to be ineffective. Rather, the code rate must be selected to transmit more information bits when the channel is favorable. As $(1 - b_{12} - b_{21})$ approaches one, the optimal value of K/N tends to one as well. In the limit, the channel behaves much like a packet erasure model: send as many bits as possible when the channel is good and ask for retransmissions whenever the message is corrupted. The data points that provide a basis for these findings are summarized in Table V.1.

V.G CONCLUSIONS

We present a methodology for the analysis and the design of digital communication systems that operate over channels with memory. The proposed approach is based on the time elapsed between the onset of the communication process and its termination. Results also extend to the asymptotic decay rates of mean service and mean sojourn time. Emphasis is on the selection of code rate for protection against erasures. We provide a simple mathematical characterization of the first-passage time to an empty queue and the large deviations on the mean service and

mean transmission time, along with a computationally efficient means to compare the performance of various implementation candidates.

The properties of coded systems are explored through a numerical study. Optimal code rates appear robust to initial buffer conditions at the transmitter. That is, the number of information bits to be sent from the source to the destination does not significantly affect the optimal operating point of the encoder. Optimal operation is achieved with very similar K values for mean first-passage times and various crossings of the cumulative distribution function.

For both mean service rate and mean sojourn time, it seems that the optimal operating point of a system in terms of code rate selection depends heavily on the needs of the underlying traffic. In particular, delay-adverse applications may perform better with coarse quantization and low-rate codes. On the other hand, delay tolerant applications may be able to use a higher rate on the same physical channel. This phenomenon is closely related to the concept of effective capacity.

Lastly, the optimal code rate depends heavily on channel memory. This suggest that, for systems with fixed block lengths, the channel parameters should be estimated and fed back to the encoder for optimal operation. This naturally leads to adaptive strategies and possibly state-aware encoding schemes at the source.

V.H APPENDIX

V.H.1 Proof of Theorem 100

We begin this proof by introducing a convenient notation for abstract sequences. Let $\{a_s\}$ be a discrete-time sequence and assume that r and t are two integers with $r < t$. We use a_r^t to denote the subsequence $a_r, a_{r+1}, \ldots, a_t$.

Suppose $u_t = (i_t, q_t) \in \mathcal{C} \times \mathbb{N}_0$ for every $t \geq 0$. Since $\{U_s\}_{s \geq 0}$ is a discrete-time stochastic process whose elements take on values in a finite set, it suffices to show that

$$\Pr\left(U_{s+1} = u_{s+1} | U_0^s = u_0^s\right) = \Pr\left(U_{s+1} = u_{s+1} | U_s = u_s\right)$$

in order to prove that this process is Markov. In general, the probability on the left hand side can be expressed as

$$\Pr\left(C_{(s+1)N+1} = i_{s+1} | U_0^s = u_0^s\right) \Pr\left(Q_{s+1} = q_{s+1} | U_0^s = u_0^s, C_{(s+1)N+1} = i_{s+1}\right).$$

We know that the state of the channel at the onset of codeword $s+1$, labeled

$C_{(s+1)N+1}$, is conditionally independent of the subsequence Q_0^s and the channel states $C_1^{(s-1)N+1}$, given C_{sN+1}. Thus, we get

$$\Pr\left(C_{(s+1)N+1} = i_{s+1} | U_0^s = u_0^s\right) = \Pr\left(C_{(s+1)N+1} = i_{s+1} | C_{sN+1} = i_s\right).$$

The length of the queue Q_{s+1} at time $s+1$ is either Q_s or $Q_s - 1$, depending on whether a codeword is successfully decoded at time s. For a non-empty queue, this depends solely on the generated codebook and the channel realizations during the transmission cycle of the codeword s. As such, we can write

$$\Pr\big(Q_{s+1} = q_{s+1} | U_0^s = u_0^s, C_{(s+1)N+1} = i_{s+1}\big) = \Pr\big(Q_{s+1} = q_{s+1} | U_s = u_s, C_{(s+1)N+1} = i_{s+1}\big).$$

Collecting these two results, we conclude that $\{U_s\}$ possesses the Markov property.

V.H.2 Proof of Proposition 102

Notice that the proposition is trivially true when $e > p$. The only case of interest then corresponds to $e \leq p$. We observe that, through a change in indexing, we can write

$$\prod_{l=0}^{n+e-1} \left(1 - 2^{l-p-n}\right) = \prod_{l=-n}^{e-1} \left(1 - 2^{l-p}\right).$$

As such, we readily see that $P_{\mathrm{f}}(p+n, e+n)$ is monotonically increasing in n. The difference between this function and $P_{\mathrm{f}}(p,e)$ is obtained as follows,

$$\begin{aligned}
P_{\mathrm{f}}(p+n, e+n) - P_{\mathrm{f}}(p,e) &= \prod_{l=0}^{e-1} \left(1 - 2^{l-p}\right) - \prod_{l=0}^{n+e-1} \left(1 - 2^{l-n-p}\right) \\
&= \prod_{l=n}^{n+e-1} \left(1 - 2^{l-n-p}\right) - \prod_{l=0}^{n+e-1} \left(1 - 2^{l-n-p}\right) \\
&= \prod_{l=n}^{n+e-1} \left(1 - 2^{l-n-p}\right) \left(1 - \prod_{l=0}^{n-1} \left(1 - 2^{l-n-p}\right)\right) \\
&\leq 1 - \prod_{l=0}^{n-1} \left(1 - 2^{l-n-p}\right) \overset{(a)}{\leq} \sum_{l=0}^{n-1} 2^{l-n-p} \\
&= \sum_{l=0}^{n-1} 2^{-l-1-p} \leq \sum_{l=0}^{\infty} 2^{-l-1-p} = 2^{-p}.
\end{aligned}$$

Step (a) follows from an n-variable version of the inequality $1-(1-p_1)(1-p_2) \leq p_1+p_2$ where $0 \leq p_1, p_2 \leq 1$. This concludes the demonstration.

V.H.3 Proof of Lemma 104

When $\lambda < -\log \varrho(\mathbf{K})$, the spectral radius of the matrix $\mathbf{K}e^{\lambda}$ is strictly less than one and, consequently, the matrix $\mathbf{I}-\mathbf{K}e^{\lambda}$ is invertible. The finiteness of $\mathbf{G}_T\left(e^{\lambda}\right)$ immediately follows. We then turn to the alternate case, which we prove by contradiction.

Assume that, for some $\lambda \geq -\log \varrho(\mathbf{K})$, matrix $\mathbf{G}_T\left(e^{\lambda}\right)$ exists over the non-negative real numbers. Note that this condition implies $\varrho(\mathbf{K}) > 0$. For convenience, we wish to work with the irreducible normal form of $\mathbf{K}$ [196]. That is, there exists a permutation matrix $\mathbf{P}$ such that

$$\tilde{\mathbf{K}} = \mathbf{P}^{\mathrm{T}}\mathbf{K}\mathbf{P} = \begin{bmatrix} \mathbf{\Psi}_1 & \mathbf{\Phi}_{12} & \cdots & \mathbf{\Phi}_{1h} \\ \mathbf{0} & \mathbf{\Psi}_2 & \cdots & \mathbf{\Phi}_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Psi}_h \end{bmatrix}$$

in which each $\mathbf{\Psi}_i$ is either irreducible or a zero matrix. Of course, this reordering also affects $\mathbf{M}$,

$$\tilde{\mathbf{M}} = \mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P}.$$

However, this transformation does not alter the spectrum of $\mathbf{K}$ or $\mathbf{M}$. We note that all the states corresponding to an irreducible $\mathbf{\Psi}_i$ belong to a same communicating class, which we denote by $\mathcal{C}_i$. Looking at the block triangular structure of $\tilde{\mathbf{K}}$, we gather that the eigenvalues of $\tilde{\mathbf{K}}$ correspond to the union of the eigenvalues of $\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_h$. Thus, there exists an integer j such that $\varrho(\mathbf{\Psi}_j) = \varrho(\mathbf{K})$.

Since matrix $\mathbf{\Psi}_j$ is non-negative and irreducible, the Perron-Frobenius theorem applies and there exists an eigenvector $\mathbf{v}$, with positive components, such that

$$\mathbf{v}\mathbf{\Psi}_j = \varrho(\mathbf{\Psi}_j)\mathbf{v} = \varrho(\mathbf{K})\mathbf{v}.$$

Without loss of generality, we can assume that $\mathbf{v}$ is normalized to one. Let $\mathbf{w}$ be a probability distribution with weight $\mathbf{v}$ over the states associated with $\mathbf{\Psi}_j$ and zero elsewhere, i.e.,

$$\mathbf{w} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{v} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

Because $\mathbf{v}$ is an eigenvector of $\mathbf{\Psi}_j$, we have

$$\mathbf{w}\left(\tilde{\mathbf{K}}e^{\lambda}\right)^t = \begin{bmatrix}\mathbf{0} & \cdots & \mathbf{0} & \left(\varrho(\mathbf{K})e^{\lambda}\right)^t\mathbf{v} & * & \cdots & *\end{bmatrix}$$

and, correspondingly,

$$\mathbf{w}\sum_{t=0}^{\infty}\tilde{\mathbf{K}}^t e^{t\lambda} = \sum_{t=0}^{\infty}\mathbf{w}\tilde{\mathbf{K}}^t e^{t\lambda} = \begin{bmatrix}\mathbf{0} & \cdots & \mathbf{0}\sum_{t=0}^{\infty}\left(\varrho(\mathbf{K})e^{\lambda}\right)^t\mathbf{v} & * & \cdots & *\end{bmatrix}.$$

We note that the multiplicative factor $\sum_{t=0}^{\infty}\left(\varrho(\mathbf{K})e^{\lambda}\right)^t$ is a divergent sum that increases to infinity. In fact, all the components of $\mathbf{w}\sum_{t=0}^{\infty}\tilde{\mathbf{K}}^t e^{t\lambda}$ corresponding to states that are accessible from $\mathcal{C}_j$ must also diverge [196]. Since by assumption the elements of

$$\tilde{\mathbf{G}}_T\left(e^{\lambda}\right) = \left(\sum_{t=0}^{\infty}\tilde{\mathbf{K}}^t e^{t\lambda}\right)\tilde{\mathbf{M}}e^{\lambda}$$

remain finite, we conclude that any state accessible from $\mathcal{C}_j$ must lie in the nullspace of $\tilde{\mathbf{M}}$. This necessarily means that $\mathbf{w}\tilde{\mathbf{G}}_T\left(e^{\lambda}\right) = \mathbf{0}$ and, consequently, $\mathbf{w}\tilde{\mathbf{G}}_T(1) = \mathbf{0}$ because $\tilde{\mathbf{K}}$ and $\tilde{\mathbf{M}}$ are non-negative matrices. In other words, we have created a valid probability distribution $\mathbf{w}$ for which $\mathbf{w}\tilde{\mathbf{G}}_T(1) = \mathbf{0}$. Equivalently, in the original domain, we can rewrite this equation as $\mathbf{w}\mathbf{P}^{\mathrm{T}}\mathbf{G}_T(1) = \mathbf{0}$. But this equation violates our assumption that T is finite almost surely. We then conclude, by contradiction, that not all entries of $\mathbf{G}_T\left(e^{\lambda}\right)$ are finite when $\lambda \geq -\log\varrho(\mathbf{K})$.

V.H.4 Proof of Corollary 105

As a straightforward application of Lemma 104, we can show that $\varrho(\mathbf{K}) < 1$. By design, we know that T is finite almost surely. Then, from the definition of the matrix generating function $\mathbf{G}_T(z)$ in (V.9), we gather that

$$[\mathbf{G}_T(1)]_{ij} = \mathrm{E}\left[\mathbf{1}_{\{C_{NT+1}=j\}}|C_1 = i\right] = \Pr\left(C_{NT+1} = j|C_1 = i\right).$$

That is, $\mathbf{G}_T(1)$ is a right stochastic matrix.

Since $\mathbf{K}$ is a substochastic matrix, we already have the relation $\varrho(\mathbf{K}) \leq 1$. We wish to show that, in the current framework, this inequality is strict. Suppose that $\varrho(\mathbf{K}) = 1$. Lemma 104 states that, if $\lambda = -\log\varrho(\mathbf{K}) = 0$, then not all entries of $\mathbf{G}_T\left(e^0\right) = \mathbf{G}_T(1)$ can be finite. In particular, $\mathbf{G}_T(1)$ cannot be a right stochastic

matrix. This leads to an obvious contradiction, which indicates that $\varrho(\mathbf{K}) < 1$, as desired.

V.H.5 Proof of Proposition 107

For the first part of this proof, we assume that $\lambda < -\log \varrho(\mathbf{K})$. The spectral radius of $\mathbf{K}e^{\lambda}$ is then strictly less than one and, as such, $\left(\mathbf{I} - \mathbf{K}e^{\lambda}\right)$ is invertible. This implies that the matrix

$$\mathbf{G}_T\left(e^{\lambda}\right) = \left(\sum_{t=0}^{\infty} \mathbf{K}^t e^{t\lambda}\right) \mathbf{M}e^{\lambda} = \left(\mathbf{I} - \mathbf{K}e^{\lambda}\right)^{-1} \mathbf{M}e^{\lambda}$$

is well-defined over the real numbers. Under Assumption 106, we know that $\mathbf{G}_T(1)$ is an irreducible matrix. This readily implies that $\mathbf{G}_T\left(e^{\lambda}\right)$ is also irreducible. We can therefore apply the Perron-Frobenius theorem [182, Th. 3.1.1], whose asymptotic properties lead directly to $\Lambda(\lambda)$.

For the second case, we suppose that $\lambda \geq -\log \varrho(\mathbf{K})$. By Lemma 104, we know that at least one entry of $\mathbf{G}_T\left(e^{\lambda}\right)$ is equal to infinity. We can use the irreducibility of this matrix to argue that each row in $\left(\mathbf{G}_T\left(e^{\lambda}\right)\right)^k$ has at least one entry that is infinite. Since π_0 is a probability distribution,

$$\mathrm{E}\left[e^{\lambda(T_1+\cdots+T_k)}\right] = \pi_0 \left(\mathbf{G}_T\left(e^{\lambda}\right)\right)^k \mathbf{1} = \infty.$$

For any $m > k$, we have

$$\begin{aligned}\Lambda_m(m\lambda) &= \log \mathrm{E}\left[e^{m\lambda Y_m}\right] = \log \mathrm{E}\left[e^{\lambda(T_1+\cdots+T_m)}\right] \\ &\geq \log \mathrm{E}\left[e^{\lambda(T_1+\cdots+T_k)}\right] = \infty.\end{aligned}$$

Consequently, whenever $\lambda \geq -\log \varrho(\mathbf{K})$, we get

$$\Lambda(\lambda) = \lim_{m\to\infty} \frac{1}{m} \Lambda_m(m\lambda) = \infty,$$

as desired.

V.H.6 Proof of Proposition 113

For the sake of completeness, we offer a brief proof for Proposition 113. As an initial step for this demonstration, we establish a few key properties. The processes

$\{Y_m\}$ and $\{Z_s\}$ converge almost surely, i.e.,

$$Y_m = \frac{1}{m}\sum_{q=1}^{m} T_q \xrightarrow{a.s.} \bar{T}$$
$$Z_s = \frac{1}{s}\sum_{t=1}^{s} D_t \xrightarrow{a.s.} \bar{D},$$

where $\bar{T}$ are $\bar{D}$ are constants. Moreover, $\bar{T}$ and $\bar{D}$ have a reciprocal relation, i.e., $\bar{T} = 1/\bar{D}$.

Recall that process $\left\{V_s = \left(C_{(s+1)N+1}, D_s\right)\right\}$ is a finite-state Markov chain with irreducible transition probability matrix $\mathbf{\Pi}$. Also, $D_s = f(V_s)$ is a (trivial) bounded function. Then, by the ergodic theorem for Markov chains [180], we have

$$\Pr\left(\lim_{s\to\infty} \frac{1}{s}\sum_{t=1}^{s} D_t = \bar{D}\right) = 1.$$

Let Ω_1 be the subset of Ω defined by

$$\Omega_1 = \left\{\omega : \frac{1}{s}\sum_{t=1}^{s} D_t(\omega) \to \bar{D}\right\}.$$

Clearly, for any $\omega \in \Omega_1$, we necessarily have

$$N(s,\omega) = \sum_{t=1}^{s} D_t(\omega) \to \infty.$$

Consider the empirical average defined by

$$\frac{1}{m}\sum_{q=1}^{m} T_q. \tag{V.20}$$

We wish to show that this sequence converges almost surely to $1/\bar{D}$ as m increases to infinity. For any $\omega \in \Omega_1$, we have

$$\sum_{q=1}^{N(s,\omega)} T_q(\omega) \le s \le \sum_{q=1}^{N(s,\omega)+1} T_q(\omega).$$

As such, we get the inequality

$$\frac{1}{N(s,\omega)} \sum_{q=1}^{N(s,\omega)} T_q(\omega) \leq \frac{s}{N(s,\omega)} \to \frac{1}{\bar{D}}.$$

In a similar fashion, we obtain

$$\begin{aligned}\frac{1}{N(s,\omega)+1} \sum_{q=1}^{N(s,\omega)+1} T_q(\omega) &\geq \frac{s}{N(s,\omega)+1} \\ &= \frac{N(s,\omega)}{N(s,\omega)+1} \frac{s}{N(s,\omega)} \to \frac{1}{\bar{D}}.\end{aligned}$$

It follows that, for any $\omega \in \Omega_1$, we get

$$\frac{1}{N(s,\omega)} \sum_{q=1}^{N(s,\omega)} T_q(\omega) \to \frac{1}{\bar{D}}. \tag{V.21}$$

To complete the proof, we must connect this result to the sequence in (V.20). We emphasize that, for any $\omega \in \Omega_1$ and for any $m \in \mathbb{N}$, there exists s such that $N(s,\omega) = m$ because $N(s,\omega)$ increases by at most one at every step. It follows that (V.20) is a subsequence of convergent sequence (V.21). They must then share the same limit. Collecting these results, we gather that

$$\Pr\left(\lim_{m\to\infty} \frac{1}{m} \sum_{q=1}^{m} T_q = \frac{1}{\bar{D}} \right) = 1.$$

As a side note, it is possible to show that

$$\begin{aligned}\bar{D} &= \mathrm{E}_{\pi_D}\left[D_t\right] = \pi_D \mathbf{M}\mathbf{1} \\ \bar{T} &= \mathrm{E}_{\pi_T}\left[T_q\right] = \pi_T \left[\lim_{\lambda\uparrow 0} \frac{d}{d\lambda} \mathbf{G}_T\left(e^{\lambda}\right) \right] \mathbf{1},\end{aligned}$$

where $\frac{d}{d\lambda}\mathbf{G}_T(e^{\lambda})$ denotes the entrywise derivative. Above, π_D and π_T represent the invariant distributions of the channel and the stochastic matrix $\mathbf{G}_T(1)$, respectively.

Our strategy to finish this proof is to establish the claimed result for rational numbers, and then invoke continuity to get a full characterization. From our hypotheses, we know that the rate functions $\Lambda^*(\cdot)$ and $I(\cdot)$ are finite in the open intervals $(1,\infty)$

and $(0,1)$, respectively. We note that these functions are also convex over these intervals and, hence, continuous. Let $r = p/q$, where $p, q \in \mathbb{N}$, be a rational number less than one. Recall that $I(\cdot)$ is convex and, therefore, continuous over $(0,1)$. Then, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\begin{aligned} -I(r) - \varepsilon &\leq \liminf_{n\to\infty} \frac{1}{np} \log \Pr\left(Z_{np} \in (r-\delta, r+\delta)\right) \\ &\leq \limsup_{n\to\infty} \frac{1}{np} \log \Pr\left(Z_{np} \in (r-\delta, r+\delta)\right) \leq -I(r) + \varepsilon. \end{aligned}$$

Taking the limit as $\delta \to 0$, we get

$$\begin{aligned} &\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{np} \log \Pr\left(Z_{np} \in (r-\delta, r+\delta)\right) \\ &= \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{np} \log \Pr\left(Z_{np} \in (r-\delta, r+\delta)\right) = -I(r). \end{aligned}$$

A similar argument applies to $\{Y_m\}$. Noting that $q/p \in (1,\infty)$, we gather that $\Lambda^*(\cdot)$ is continuous in a neighborhood of $1/r$. Then, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\begin{aligned} -\Lambda^*\left(\frac{1}{r}\right) - \varepsilon &\leq \liminf_{n\to\infty} \frac{1}{nq} \log \Pr\left(Y_{nq} \in \left(\frac{1}{r} - \delta, \frac{1}{r} + \delta\right)\right) \\ &\leq \limsup_{n\to\infty} \frac{1}{nq} \log \Pr\left(Y_{nq} \in \left(\frac{1}{r} - \delta, \frac{1}{r} + \delta\right)\right) \\ &\leq -\Lambda^*\left(\frac{1}{r}\right) + \varepsilon. \end{aligned}$$

As before, this implies that

$$\begin{aligned} &\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{nq} \log \Pr\left(Y_{nq} \in \left(\frac{1}{r} - \delta, \frac{1}{r} + \delta\right)\right) \\ &= \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{nq} \log \Pr\left(Y_{nq} \in \left(\frac{1}{r} - \delta, \frac{1}{r} + \delta\right)\right) \\ &= -\Lambda^*\left(\frac{1}{r}\right). \end{aligned}$$

We stress that the rate functions $\Lambda^*(\cdot)$ and $I(\cdot)$ vanish at $\bar{T}$ and $\bar{D}$, respectively.

At this point, we need to consider two separate cases. First, suppose $r < \bar{D}$. We know that $I(\cdot)$ is a non-increasing function over interval $\left[0, \bar{D}\right)$ (see, e.g., [182, Lemma

2.2.5]). Also, in an analogous manner, rate function $\Lambda^*(\cdot)$ is non-decreasing over $\left(\bar{T}, \infty\right)$. Leveraging (V.19), we can write

$$\Pr\left(\frac{T_1 + \cdots + T_{pn}}{pn} > \frac{q}{p}\right) = \Pr\left(\frac{D_1 + \cdots + D_{qn}}{qn} < \frac{p}{q}\right).$$

By letting n go to infinity, we obtain

$$\inf_{x \in \left[\frac{1}{r}, \infty\right)} r\Lambda^*(x) = \inf_{x \in (0, r]} I(x).$$

Using the monotonic properties of these rate functions over the prescribed intervals, we get

$$r\Lambda^*\left(\frac{1}{r}\right) = \inf_{x \in \left[\frac{1}{r}, \infty\right)} r\Lambda^*(x) = \inf_{x \in (0, r]} I(x) = I(r),$$

as desired.

For the second case, assume $r > \bar{D}$. Under this constraint, the monotonic properties of the rate functions are reversed. That is, $I(\cdot)$ is non-decreasing over $\left(\bar{D}, 1\right)$ and $\Lambda^*(\cdot)$ is non-increasing over $\left(0, \bar{T}\right)$. Using these relations and the set equalities

$$\Pr\left(\frac{T_1 + \cdots + T_{pn}}{pn} < \frac{q}{p}\right) = \Pr\left(\frac{D_1 + \cdots + D_{qn}}{qn} > \frac{p}{q}\right),$$

we can write

$$r\Lambda^*\left(\frac{1}{r}\right) = \inf_{x \in \left(0, \frac{1}{r}\right]} r\Lambda^*(x) = \inf_{x \in [r, \infty)} I(x) = I(r).$$

Collecting these results, we deduce that $I(x) = x\Lambda^*\left(\frac{1}{x}\right)$ whenever $x \in \mathbb{Q} \cap (0, 1)$. Since the rational numbers are dense in $(0, 1)$ and the two rate functions are continuous, this equality must also hold for any real number in $(0, 1)$.

CHAPTER VI

CONCLUSION & FUTURE RESEARCH

While coding theory has come a long way since the seminal work of Shannon and dramatically shaped the current landscape of telecommunications, we believe there are several interesting phenomena to discover that will further shape the future. Below, we list some open questions that emanate from this dissertation.

- Even though the threshold saturation in spatially-coupled codes is a universal phenomenon, the biggest hurdle in its adoption in current communication systems is the rate loss due to boundary. Mitigating this necessitates large chain length and consequently large blocklengths. One question in this direction is to quantify the boundary information required to kick-start the "encoding wave". It is also interesting to investigate the potential of threshold saturation in quantum systems.

- We have seen the utility of the BPGD algorithm in Chapter III in the context of source coding, where the standard BP algorithms fall far short of the desired performance. Also, the iterative quantization procedure in Algorithm 2 cannot constrain the weight of the output sequence. It would be useful to see whether one can utilize the BPGD algorithm to accomplish this and hence achieve the capacity region of the multi-write WOM system. More generally, it remains to see whether this algorithm plays a crucial role in other graphical models, where standard BP algorithms suffer from convergence.

- Although the study of Reed-Muller codes in Chapter IV has been a theoretical curiosity, it opens up a plethora of questions. The most important question that warrants attention here is the generality of the capacity achievability of structured codes such as Reed-Muller and BCH codes. More precisely, do Reed-Muller and BCH codes achieve capacity on general channels? Also, the early popularity of Reed-Muller lends to the availability of low complexity decoding algorithms. Unfortunately, these algorithms do not scale well for large blocklengths and arbitrary rates. It is important to explore algorithmic aspects of these codes in these regimes.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell Syst. Techn. J.*, vol. 27, pp. 379–423, 623–656, July / Oct. 1948.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. IEEE Int. Conf. Commun.*, vol. 2. Geneva, Switzerland: IEEE, May 1993, pp. 1064–1070.

[3] D. Spielman, "Linear-time encodable and decodable error-correcting codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1723–1731, Nov 1996.

[4] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electronic Letters*, vol. 32, no. 18, pp. 1645–1646, Aug. 1996.

[5] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA, USA: The M.I.T. Press, 1963.

[6] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[7] J. Felstrom and K. S. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 2181–2191, Sept. 1999.

[8] A. Sridharan, M. Lentmaier, D. J. Costello, and K. S. Zigangirov, "Convergence analysis of a class of LDPC convolutional codes for the erasure channel," in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, USA, Oct. 2004, pp. 953–962.

[9] M. Lentmaier, A. Sridharan, K. S. Zigangirov, and D. J. Costello, "Terminated LDPC convolutional codes with thresholds close to capacity," in *Proc. IEEE Int. Symp. Inform. Theory*, Adelaide, Australia, Sept. 2005, pp. 1372–1376.

[10] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.

[11] A. Yedla, Y.-Y. Jian, P. S. Nguyen, and H. D. Pfister, "A simple proof of threshold saturation for coupled scalar recursions," in *Proc. Int. Symp. on Turbo Codes & Iterative Inform. Proc.*, Aug. 2012, pp. 51–55.

[12] ——, "A simple proof of threshold saturation for coupled vector recursions," in *Proc. IEEE Inform. Theory Workshop*, Sept. 2012, pp. 25–29.

[13] M. J. Wainwright and E. Martinian, "Low-density graph codes that are optimal for binning and coding with side information," *IEEE Trans. Inform. Theory*, vol. 55, no. 3, pp. 1061–1079, March 2009.

[14] E. Abbe, A. Shpilka, and A. Wigderson, "Reed-muller codes for random erasures and errors," in *Proc. of the Annual ACM Symp. on Theory of Comp.*, ser. STOC '15. New York, NY, USA: ACM, 2015, pp. 297–306.

[15] F. J. MacWilliams and N. J. A. Sloane, *The theory of error correcting codes*. Elsevier, 1977, vol. 16.

[16] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2004, iSBN-13: 978-0130426727.

[17] P. Delsarte, J. Goethals, and F. M. Williams, "On generalized Reed-Muller codes and their relatives," *Inform. and Control*, vol. 16, no. 5, pp. 403–442, 1970.

[18] M. Lentmaier, A. Sridharan, D. J. Costello, and K. S. Zigangirov, "Iterative decoding threshold analysis for LDPC convolutional codes," *IEEE Trans. Inform. Theory*, vol. 56, no. 10, pp. 5274–5289, Oct. 2010.

[19] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Trans. Inform. Theory*, vol. 59, no. 12, pp. 7761–7813, Dec. 2013.

[20] S. Kudekar, C. Méasson, T. Richardson, and R. Urbanke, "Threshold saturation on BMS channels via spatial coupling," in *Proc. Int. Symp. on Turbo Codes & Iterative Inform. Proc.*, Sept. 2010, pp. 309–313.

[21] V. Rathi, R. Urbanke, M. Andersson, and M. Skoglund, "Rate-equi-vocation optimally spatially coupled LDPC codes for the BEC wiretap channel," in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, July 2011, pp. 2393–2397.

[22] A. Yedla, H. D. Pfister, and K. R. Narayanan, "Universality for the noisy Slepian-Wolf problem via spatial coupling," in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, July 2011, pp. 2567–2571.

[23] S. Kudekar and K. Kasai, "Threshold saturation on channels with memory via spatial coupling," in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, July 2011, pp. 2562–2566.

[24] ——, "Spatially coupled codes over the multiple access channel," in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, July 2011, pp. 2816–2820.

[25] P. S. Nguyen, A. Yedla, H. D. Pfister, and K. R. Narayanan, "Spatially-coupled codes and threshold saturation on intersymbol-interference channels," Oct. 2011, [Online]. Available: http://arxiv.org/abs/1107.3253.

[26] K. Takeuchi, T. Tanaka, and T. Kawabata, "Improvement of BP-based CDMA multiuser detection by spatial coupling," in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, July 2011, pp. 1489–1493.

[27] C. Schlegel and D. Truhachev, "Multiple access demodulation in the lifted signal graph with spatial coupling," in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, July 2011, pp. 2989–2993.

[28] Y.-Y. Jian, H. D. Pfister, and K. R. Narayanan, "Approaching capacity at high rates with iterative hard-decision decoding," in *Proc. IEEE Int. Symp. Inform. Theory*, July 2012, pp. 2696–2700.

[29] S. H. Hassani, N. Macris, and R. Urbanke, "Coupled graphical models and their thresholds," in *Proc. IEEE Inform. Theory Workshop*, Dublin, Ireland, Jan. 2010, pp. 1–5.

[30] ——, "Chains of mean-field models," *J. Stat. Mech.*, p. P02011, 2012.

[31] ——, "Threshold saturation in spatially coupled constraint satisfaction problems," *J. Stat. Phys.*, vol. 150, no. 5, pp. 807–850, 2013.

[32] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, USA, Oct. 2010, pp. 347–353.

[33] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, p. 021005, May 2012.

[34] D. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7434–7464, Nov. 2013.

[35] K. Takeuchi, T. Tanaka, and T. Kawabata, "A phenomenological study on threshold improvement via spatial coupling," *IEICE Trans. Fundamentals*, vol. E95-A, no. 5, pp. 974–977, 2012.

[36] N. Macris, "Griffith–Kelly–Sherman correlation inequalities: A useful tool in the theory of error correcting codes," *IEEE Trans. Inform. Theory*, vol. 53, no. 2, pp. 664–683, 2007.

[37] R. Mori, "Connection between annealed free energy and belief propagation on random factor graph ensembles," in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, July 2011, pp. 2010–2014.

[38] T. J. Richardson and R. L. Urbanke, *Modern Coding Theory*. New York, NY: Cambridge University Press, 2008.

[39] A. Montanari, "Tight bounds for LDPC and LDGM codes under MAP decoding," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3221–3246, Sept. 2005.

[40] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, "A proof of threshold saturation for spatially-coupled LDPC codes on BMS channels," in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, USA, Oct. 2012, pp. 176–184.

[41] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[42] S. Kudekar and N. Macris, "Sharp bounds for optimal decoding of low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 55, no. 10, pp. 4635–4650, Oct. 2009.

[43] A. Giurgiu, N. Macris, and R. Urbanke, "Spatial coupling as a proof technique," Jan. 2013, [Online]. Available: http://arxiv.org/abs/1301.5676.

[44] C. Méasson, A. Montanari, T. J. Richardson, and R. Urbanke, "The generalized area theorem and some of its consequences," *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 4793–4821, Nov. 2009.

[45] M. Luby, "LT codes," in *Proc. of the 43rd Symp. on Foundations of Comp. Sci.*, Washington, D.C., USA, June 2002, p. 271.

[46] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2551–2567, 2006.

[47] M. Wainwright, E. Maneva, and E. Martinian, "Lossy source compression using low-density generator matrix codes: Analysis and algorithms," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 1351–1368, 2010.

[48] V. Aref, N. Macris, and M. Vuffray, "Approaching the rate-distortion limit with spatial coupling, belief propagation, and decimation," *IEEE Trans. Inform. Theory*, vol. 61, no. 7, pp. 3954–3979, July 2015.

[49] V. Aref and R. Urbanke, "Universal rateless codes from coupled LT codes," in *Proc. IEEE Inform. Theory Workshop*, 2011, pp. 277–281.

[50] W. Feller, *An Introduction to Probability Theory and Its Applications*. Wiley, 1971, vol. 2.

[51] G. Folland, *Real analysis: modern techniques and their applications*. Wiley, 1999.

[52] F. Steutel and K. van Harn, *Infinite Divisibility of Probability Distributions on the Real Line*, ser. Chapman & Hall/CRC Pure and Applied Mathematics. Taylor & Francis, 2003.

[53] M. Mezard and A. Montanari, *Information, Physics, and Computation*. New York, NY: Oxford University Press, 2009.

[54] J. M. Walsh and P. A. Regalia, "On the relationship between belief propagation decoding and joint maximum likelihood detection," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2753–2758, Oct. 2010.

[55] K. Y. M. Wong and D. Sherrington, "Graph bipartitioning and spin glasses on a random network of fixed finite valence," *J. Phys. A: Mathematical and General*, vol. 20, no. 12, p. L793, 1987.

[56] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[57] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Inform. Transm.*, vol. 9, pp. 19–31, 1980.

[58] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., ser. Wiley Series in Telecommunications. Wiley, 2006.

[59] R. L. Rivest and A. Shamir, "How to reuse a write-once memory," *Information and Control*, vol. 55, no. 13, pp. 1 – 19, 1982.

[60] C. Heegard, "On the capacity of permanent memory," *IEEE Trans. Inf. Theory*, vol. 31, no. 1, pp. 34–42, Jan 1985.

[61] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," in *Proc. Asilomar Conf. on Signals, Systems & Computers*, vol. 1, 2003, pp. 825–829.

[62] A. Liveris, Z. Xiong, and C. Georghiades, "Nested convolutional/turbo codes for the binary Wyner-Ziv problem," in *Proc. Intl. Conf. on Image Proc.*, vol. 1, 2003, pp. I–601.

[63] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3417–3432, 2005.

[64] Y. Sun, Y. Yang, A. Liveris, V. Stankovic, and Z. Xiong, "Near-capacity dirty-paper code design: A source-channel coding approach," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3013–3031, 2009.

[65] J. K. Wolf, A. D. Wyner, J. Ziv, and J. Krner, "Coding for a write-once memory," *AT&T Bell Laboratories Technical Journal*, vol. 63, no. 6, pp. 1089–1112, 1984.

[66] F. Merkx, "Womcodes constructed with projective geometries," *Traitement du signal*, vol. 1, pp. 227–231, 1984.

[67] G. Cohen, P. Godlewski, and F. Merkx, "Linear binary code for write-once memories (corresp.)," *IEEE Trans. Inf. Theory*, vol. 32, no. 5, pp. 697–700, Sep 1986.

[68] Y. Wu, "Low complexity codes for writing a write-once memory twice," in *IEEE Int. Symp. Inf. Th.*, June 2010, pp. 1928–1932.

[69] Y. Wu and A. Jiang, "Position modulation code for rewriting write-once memories," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3692–3697, June 2011.

[70] E. Yaakobi, S. Kayser, P. Siegel, A. Vardy, and J. Wolf, "Codes for write-once memories," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5985–5999, Sept 2012.

[71] A. Shpilka, "New constructions of wom codes using the Wozencraft ensemble," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4520–4529, July 2013.

[72] D. Burshtein and A. Strugatski, "Polar write once memory codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 5088–5101, Aug 2013.

[73] E. E. Gad, W. Huang, Y. Li, and J. Bruck, "Rewriting flash memories by message passing," in *IEEE Int. Symp. on Inf. Th.*, Hong Kong, China, June 2015.

[74] G. Zémor, "Problemes combinatoires lies a l'ecriture sur des memoires," Ph.D. dissertation, ENST, Paris, France, 1989.

[75] G. Zémor and G. Cohen, "Error-correcting wom-codes," *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp. 730–734, May 1991.

[76] E. Yaakobi, P. Siegel, A. Vardy, and J. Wolf, "Multiple error-correcting wom-codes," *Information Theory, IEEE Transactions on*, vol. 58, no. 4, pp. 2220–2230, April 2012.

[77] A. Jiang, Y. Li, E. Gad, M. Langberg, and J. Bruck, "Joint rewriting and error correction in write-once memories," in *IEEE Int. Symp. Inf. Theory*, July 2013, pp. 1067–1071.

[78] E. E. Gad, Y. Li, J. Kliewer, M. Langberg, A. Jiang, and J. Bruck, "Asymmetric error correction and flash-memory rewriting using polar codes," *CoRR*, vol. abs/1410.3542, 2014.

[79] T. Murayama, "Thouless-Anderson-Palmer approach for lossy compression," *Physical Review E*, vol. 69, no. 3, p. 035105, 2004.

[80] P. Regalia, "A modified belief propagation algorithm for code word quantization," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3513–3517, 2009.

[81] S. Ciliberti, M. Mézard, and R. Zecchina, "Lossy data compression with random gates," *Physical review letters*, vol. 95, no. 3, p. 38701, 2005.

[82] M. Wainwright, E. Maneva, and E. Martinian, "Lossy source compression using low-density generator matrix codes: Analysis and algorithms," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 1351–1368, 2010.

[83] V. Aref, N. Macris, R. Urbanke, and M. Vuffray, "Lossy source coding via spatially coupled LDGM ensembles," in *Proc. IEEE Int. Symp. Inform. Theory*, Cambridge, MA, USA, July 2012, pp. 373–377.

[84] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian, "Solving constraint satisfaction problems through belief propagation-guided decimation," in *Proc. 45th Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, USA, Sept. 2007.

[85] C. H. Hsu and A. Anastasopoulos, "Capacity-achieving codes with bounded graphical complexity and maximum likelihood decoding," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 992–1006, 2010.

[86] K. Kasai and K. Sakaniwa, "Spatially-coupled MacKay-Neal codes and Hsu-Anastasopoulos codes," *IEICE Trans. Fundamentals*, vol. E94-A, no. 11, pp. 2161–2168, 2011.

[87] N. Obata, Y.-Y. Jian, K. Kasai, and H. D. Pfister, "Spatially-coupled multi-edge type LDPC codes with bounded degrees that achieve capacity on the BEC under BP decoding," in *Proc. IEEE Int. Symp. Inform. Theory*, July 2013, pp. 2433–2437.

[88] S. B. Korada and R. L. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1751–1768, 2010.

[89] E. Martinian and J. S. Yedidia, "Iterative quantization using codes on graphs," in *in Proc. Allerton Conf. Comm., Cont., and Comp.*, 2003.

[90] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, "Threshold saturation for spatially-coupled LDPC and LDGM codes on BMS channels," *IEEE Trans. Inform. Theory*, vol. 60, no. 12, pp. 7389–7415, Dec. 2014.

[91] K. Tazoe, K. Kasai, and K. Sakaniwa, "Efficient termination of spatially-coupled codes," in *IEEE Inf. Th. Workshop (ITW)*, Sept 2012, pp. 30–34.

[92] A. Iyengar, P. Siegel, R. Urbanke, and J. Wolf, "Windowed decoding of spatially coupled codes," *IEEE Tran. on Inf. Th.*, vol. 59, no. 4, pp. 2277–2292, April 2013.

[93] P. Olmos and R. Urbanke, "A scaling law to predict the finite-length performance of spatially-coupled ldpc codes," *IEEE Tran. on Inf. Th.*, vol. 61, no. 6, pp. 3164–3184, June 2015.

[94] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 399–431, March 1999.

[95] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 569–584, Feb. 2001.

[96] T. Kasami, S. Lin, and W. Peterson, "New generalizations of the Reed-Muller codes–I: Primitive codes," *IEEE Trans. Inform. Theory*, vol. 14, no. 2, pp. 189–199, Mar 1968.

[97] T. Kasami, S. Lin, and W. W. Peterson, "Some results on cyclic codes which are invariant under the affine group and their applications," *Inform. and Control*, vol. 11, no. 5, pp. 475–496, 1968.

[98] R. Ahlswede and G. Dueck, "Good codes can be produced by a few permutations," *IEEE Trans. Inform. Theory*, vol. 28, no. 3, pp. 430–443, May 1982.

[99] J. Coffey and R. Goodman, "Any code of which we cannot think is good," *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1453–1461, Nov 1990.

[100] D. J. Costello, Jr. and G. D. Forney, Jr., "Channel coding: The road to channel capacity," *Proc. of the IEEE*, vol. 95, no. 6, pp. 1150–1177, June 2007.

[101] D. Muller, "Application of Boolean algebra to switching circuit design and to error detection," *IRE Tran. on Electronic Computers*, vol. EC-3, no. 3, pp. 6–12, Sept 1954.

[102] I. Reed, "A class of multiple-error-correcting codes and the decoding scheme," *IRE Tran. on Information Theory*, vol. 4, no. 4, pp. 38–49, September 1954.

[103] I. Dumer and P. G. Farrell, "Erasure correction performance of linear block codes," in *Algebraic Coding*. Springer, 1994, pp. 316–326.

[104] C. Carlet and P. Gaborit, "On the construction of balanced boolean functions with a good algebraic immunity," in *Proc. IEEE Int. Symp. Inform. Theory*, Sept 2005, pp. 1101–1105.

[105] F. Didier, "A new upper bound on the block error probability after decoding over the erasure channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4496–4503, Oct 2006.

[106] E. Arikan, "A survey of reed-muller codes from polar coding perspective," in *Proc. IEEE Inform. Theory Workshop*, Jan 2010, pp. 1–5.

[107] M. Mondelli, S. Hassani, and R. Urbanke, "From polar to Reed-Muller codes: A technique to improve the finite-length performance," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3084–3091, Sept 2014.

[108] N. Sloane and E. Berlekamp, "Weight enumerator for second-order Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 16, no. 6, pp. 745–751, Nov 1970.

[109] T. Kasami and N. Tokura, "On the weight structure of Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 16, no. 6, pp. 752–759, Nov 1970.

[110] T. Kasami, N. Tokura, and S. Azumi, "On the weight enumeration of weights less than 2.5d of Reed-Muller codes," *Inform. and Control*, vol. 30, no. 4, pp. 380 – 395, 1976.

[111] T. Kaufman, S. Lovett, and E. Porat, "Weight distribution and list-decoding size of Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 58, no. 5, pp. 2689–2696, May 2012.

[112] V. M. Sidel'nikov and A. Pershakov, "Decoding of Reed-Muller codes with a large number of errors," *Problems of Inform. Transm.*, vol. 28, no. 3, pp. 80–94, 1992.

[113] I. Dumer, "Recursive decoding and its performance for low-rate Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 811–823, May 2004.

[114] ——, "Soft-decision decoding of reed-muller codes: a simplified algorithm," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 954–963, March 2006.

[115] I. Dumer and K. Shabunov, "Soft-decision decoding of Reed-Muller codes: recursive lists," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1260–1266, March 2006.

[116] R. Saptharishi, A. Shpilka, and B. L. Volk, "Decoding high rate Reed-Muller codes from random errors in near linear time," 2015, [Online]. Available: http://arxiv.org/abs/1503.09092.

[117] E. Arikan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Commun. Letters*, vol. 12, no. 6, pp. 447–449, June 2008.

[118] P. Camion, C. Carlet, P. Charpin, and N. Sendrier, "On correlation-immune functions," in *Advances in Cryptology–CRYPTO91*. Springer, 1992, pp. 86–100.

[119] A. Ta-Shma, D. Zuckerman, and S. Safra, "Extractors from Reed-Muller codes," in *Proc. IEEE Symp. on the Found. of Comp. Sci.* IEEE, 2001, pp. 638–647.

[120] R. Shaltiel and C. Umans, "Simple extractors for all min-entropies and a new pseudo-random generator," in *Proc. IEEE Symp. on the Found. of Comp. Sci.* IEEE, 2001, pp. 648–657.

[121] A. Canteaut, C. Carlet, P. Charpin, and C. Fontaine, "On cryptographic properties of the cosets of R(1, m)," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1494–1513, 2001.

[122] C. Carlet, D. K. Dalai, K. C. Gupta, and S. Maitra, "Algebraic immunity for cryptographically significant boolean functions: analysis and construction," *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3105–3121, 2006.

[123] F. Didier and J.-P. Tillich, "Computing the algebraic immunity efficiently," in *Fast Software Encryption*. Springer, 2006, pp. 359–374.

[124] B. Gérard and J.-P. Tillich, "Using tools from error correcting theory in linear cryptanalysis," *Adv. Linear Cryptanalysis of Block and Stream Ciphers*, vol. 7, p. 87, 2011.

[125] S. Yekhanin, "Locally decodable codes," *Found. Trends Theor. Comput. Sci.*, vol. 7, no. 4, pp. 169–174, 1992.

[126] P. Gemmell, R. Lipton, R. Rubinfeld, M. Sudan, and A. Wigderson, "Self-testing/correcting for polynomials and for approximate functions," in *STOC*, vol. 91. Citeseer, 1991, pp. 32–42.

[127] P. Gemmell and M. Sudan, "Highly resilient correctors for polynomials," *Information processing letters*, vol. 43, no. 4, pp. 169–174, 1992.

[128] T. Kaufman and M. Viderman, "Locally testable vs. locally decodable codes," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2010, pp. 670–682.

[129] E. Grigorescu, T. Kaufman, and M. Sudan, "2-transitivity is insufficient for local testability," in *Annual IEEE Conf. on Comp. Complex.*, June 2008, pp. 259–267.

[130] S. ten Brink, "Convergence of iterative decoding," *Electronic Letters*, vol. 35, no. 10, pp. 806–808, May 1999.

[131] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2657–2674, Nov. 2004.

[132] C. Méasson, A. Montanari, and R. L. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5277–5307, Dec. 2008.

[133] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.

[134] G. Kalai and S. Safra, "Threshold phenomena and influence with some perspectives from mathematics, computer science, and economics," *Comp. Complexity and Stat. Phy., Santa Fe Institute Studies in Sci. of Complexity*, vol. 19517738, 2005.

[135] G. A. Margulis, "Probabilistic characteristics of graphs with large connectivity," *Problems of Inform. Transm.*, vol. 10, no. 2, pp. 101–108, 1974.

[136] L. Russo, "An approximate zero-one law," *Prob. Th. and Related Fields*, vol. 61, no. 1, pp. 129–139, 1982.

[137] M. Talagrand, "Isoperimetry, logarithmic sobolev inequalities on the discrete cube, and margulis' graph connectivity theorem," *Geometric & Functional Analysis*, vol. 3, no. 3, pp. 295–314, 1993.

[138] ——, "On Russo's approximate zero-one law," *The Ann. of Prob.*, pp. 1576–1587, 1994.

[139] E. Friedgut and G. Kalai, "Every monotone graph property has a sharp threshold," *Proc. Amer. Math. Soc.*, vol. 124, no. 10, pp. 2993–3002, 1996.

[140] E. Friedgut and J. Bourgain, "Sharp thresholds of graph properties, and the k-sat problem," *J. Amer. Math. Soc.*, vol. 12, no. 4, pp. 1017–1054, 1999.

[141] I. Dinur and S. Safra, "On the hardness of approximating minimum vertex cover," *Ann. of Math.*, pp. 439–485, 2005.

[142] G. Zémor, "Threshold effects in codes," in *Algebraic Coding*. Springer, 1994, pp. 278–286.

[143] J.-P. Tillich and G. Zémor, "Discrete isoperimetric inequalities and the probability of a decoding error," *Combinatorics, Probability and Computing*, vol. 9, no. 05, pp. 465–479, 2000.

[144] J. Tillich and G. Zemor, "The Gaussian isoperimetric inequality and decoding error probabilities for the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 50, no. 2, pp. 328–331, Feb 2004.

[145] S. Kumar and H. D. Pfister, "Reed-Muller codes achieve capacity on erasure channels," 2015, available: http://arxiv.org/abs/1505.05123v1.

[146] S. Kudekar, M. Mondelli, E. Şaşoğlu, and R. Urbanke, "Reed-Muller codes achieve capacity on the binary erasure channel under MAP decoding," 2015, [Online]. Available: http://arxiv.org/abs/1505.05831v1.

[147] W. C. Huffman and V. Pless, *Fundamentals of error-correcting codes*. Cambridge University Press, 2003.

[148] T. Berger and P. Charpin, "The permutation group of affine-invariant extended cyclic codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 2194–2209, Nov 1996.

[149] D. Achlioptas, A. Naor, and Y. Peres, "Rigorous location of phase transitions in hard optimization problems," *Nature*, vol. 435, no. 7043, pp. 759–764, 2005.

[150] A. Coja-Oghlan, "The asymptotic k-SAT threshold," in *Proc. of the Annual ACM Symp. on Theory of Comp.*, ser. STOC '14. ACM, 2014, pp. 804–813.

[151] J. Ding, A. Sly, and N. Sun, "Proof of the satisfiability conjecture for large k," in *Proc. of the Annual ACM Symp. on Theory of Comp.*, ser. STOC '15. New York, NY, USA: ACM, 2015, pp. 59–68.

[152] R. Rossignol, "Threshold for monotone symmetric properties through a logarithmic Sobolev inequality," *The Ann. of Prob.*, vol. 34, no. 5, pp. 1707–1725, 09 2006.

[153] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson, and N. Linial, "The influence of variables in product spaces," *Israel Journal of Mathematics*, vol. 77, no. 1-2, pp. 55–64, 1992.

[154] M. Ben-Or and N. Linial, "Collective coin flipping," *Randomness and Computation*, vol. 5, pp. 91–115, 1990.

[155] J. Kahn, G. Kalai, and N. Linial, "The influence of variables on boolean functions," in *Proc. IEEE Symp. on the Found. of Comp. Sci.*, Oct 1988, pp. 68–80.

[156] J. Bourgain and G. Kalai, "Influences of variables and threshold intervals under group symmetries," *Geometric & Functional Analysis*, vol. 7, no. 3, pp. 438–461, 1997.

[157] T. Kasami, S. Lin, and W. Peterson, "Polynomial codes," *IEEE Trans. Inform. Theory*, vol. 14, no. 6, pp. 807–814, Nov 1968.

[158] P. Delsarte, "On cyclic codes that are invariant under the general linear group," *IEEE Trans. Inform. Theory*, vol. 16, no. 6, pp. 760–769, Nov 1970.

[159] T. Berger and P. Charpin, "The automorphism groups of BCH codes and of some affine-invariant codes over extension fields," *Designs, Codes and Cryptography*, vol. 18, no. 1-3, pp. 29–53, 1999.

[160] O. Ordentlich and U. Erez, "Cyclic-coded integer-forcing equalization," *IEEE Trans. Inform. Theory*, vol. 58, no. 9, pp. 5804–5815, 2012.

[161] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.

[162] R. Negi and J. M. Cioffi, "Delay-constrained capacity with causal feedback," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2478–2494, September 2002.

[163] W. Turin and M. Zorzi, "Performance analysis of delay-constrained communications over slow Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 801–807, October 2002.

[164] I. Bettesh and S. Shamai, "Optimal power and rate control for minimal average delay: The single-user case," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4115–4141, September 2006.

[165] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.

[166] S. Kittipiyakul, P. Elia, and T. Javidi, "High-SNR analysis of outage-limited communications with bursty and delay-limited information," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 746–763, February 2009.

[167] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. Autom. Control*, vol. 39, no. 5, pp. 913–931, May 1994.

[168] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, July 2003.

[169] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, November 2009.

[170] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Dispersion of the Gilbert-Elliott channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1829–1848, April 2011.

[171] Q. Liu, S. Zhou, and G. B. Giannakis, "Queuing with adaptive modulation and coding over wireless links: cross-layer analysis and design," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1142–1153, May 2005.

[172] X. Wang, Q. Liu, and G. B. Giannakis, "Analyzing and optimizing adaptive modulation coding jointly with ARQ for QoS-guaranteed traffic," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 710–720, March 2007.

[173] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002.

[174] R. Negi and S. Goel, "An information-theoretic approach to queuing in wireless channels with large delay bounds," in *IEEE Global Telecomm. Conf.*, December 2004, pp. 116–122.

[175] S. Goel and R. Negi, "The queued-code in finite-state Markov fading channels with large delay bounds," in *IEEE Int. Symp. Inf. Theory*, July 2006, pp. 30–34.

[176] R. Fantacci, "Queuing analysis of the selective repeat automatic repeat request protocol wireless packet networks," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 258–264, May 1996.

[177] R. E. Azouzi and E. Altman, "A queuing analysis of packet dropping over a wireless link with retransmissions," in *Personal Wireless Commun.* Springer Berlin / Heidelberg, 2003, pp. 321–333.

[178] H. J. Kushner, *Heavy Traffic Analysis of Controlled Queueing and Communications Networks*. Springer, 2001.

[179] W. Wu, A. Arapostathis, and S. Shakkottai, "Optimal power allocation for a time-varying wireless channel under heavy-traffic approximation," *IEEE Trans. Autom. Control*, vol. 51, no. 4, pp. 580–594, April 2006.

[180] J. R. Norris, *Markov Chains*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[181] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. Addison-Wesley, 1994.

[182] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications,* 2nd ed. Springer Verlag, 2009, vol. 38.

[183] L. Kleinrock, *Queueing Systems. Volume 1: Theory.* Wiley-Interscience, 1975.

[184] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, *Fundamentals of Queueing Theory*, 4th ed., ser. Probability and Statistics. Wiley, 2008.

[185] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, no. 9, pp. 1253–1265, 1960.

[186] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, no. 9, pp. 1977–1997, 1963.

[187] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley-Interscience, 1991.

[188] H. S. Wang and N. Moayeri, "Finite state Markov channel – A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, February 1995.

[189] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, November 1999.

[190] P. Sadeghi, R. A. Kennedy, P. B. Rapajic, and R. Shams, "Finite-state Markov modeling of fading channels: A survey of principles and applications," *IEEE Signal Process. Mag.*, vol. 25, no. 5, pp. 57–80, September 2008.

[191] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[192] L. Wilhelmsson and L. B. Milstein, "On the effect of imperfect interleaving for the Gilbert-Elliott channel," *IEEE Trans. Commun.*, vol. 47, no. 5, pp. 681–688, May 1999.

[193] R. A. Comroe and D. J. Costello Jr., "ARQ schemes for data transmission in mobile radio systems," *IEEE Trans. Commun.*, vol. 2, no. 4, pp. 472–481, July 1984.

[194] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes," *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1311–1321, August 2004.

[195] L. B. Le, E. Hossain, and M. Zorzi, "Queueing analysis for GBN and SR ARQ protocols under dynamic radio link adaptation with non-zero feedback delay," *IEEE Trans. Wireless Commun.*, vol. 6, no. 9, pp. 3418–3428, September 2007.

[196] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[197] P. Lancaster and M. Tismenetsky, *The theory of matrices: with applications*, 2nd ed. Academic Press, 1985.

[198] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. New York, NY, USA: Cambridge University Press, 2005.